

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2004, 01:28:51 ; Search time 154 Seconds

(without alignments)  
444.918 Million cell updates/sec

Title: US-09-084-691b-206

Perfect score: 1045

Sequence: 1 MSTLPKPKTKNTNRRPT.....CSFSIFLLALLSCLTTPASA 191

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq 23Sep04:\*

- 1: geneseqp1980s:\*
- 2: geneseqp1990s:\*
- 3: geneseqp2000s:\*
- 4: geneseqp2001s:\*
- 5: geneseqp2002s:\*
- 6: geneseqp2003as:\*
- 7: geneseqp2003bs:\*
- 8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1045	100.0	191	2	AAR92987 Hepatitis
2	996	95.3	191	2	AAR92972 Hepatitis
3	982	94.0	191	2	AAR92974 Hepatitis
4	982	94.0	191	2	AAR92973 Hepatitis
5	981	93.9	191	2	AAR92953 Hepatitis
6	980	93.8	191	2	AAR92978 Hepatitis
7	978	93.6	191	2	AAR92976 Hepatitis
8	977	93.5	191	2	AAR92977 Hepatitis
9	977	93.5	196	2	AAR74048 Synthetic
10	977	93.5	196	2	AAR74047 Synthetic
11	977	93.5	319	2	AAR96546 Hepatitis
12	977	93.5	326	2	AAR922137 HCV-HC59
13	977	93.5	733	2	AAR38278 NAMB hepa
14	977	93.5	2894	2	AAR24440 Composite
15	977	93.5	2894	2	AAR70230 Composite
16	977	93.5	3011	2	AAR66995 Hepatitis
17	976	93.4	191	2	AAR44010 Hepatitis
18	976	93.4	191	2	AAR92938 Hepatitis
19	976	93.4	191	2	AAR92941 Hepatitis
20	976	93.4	191	2	AAR92937 Hepatitis
21	976	93.4	191	2	AAR92939 Hepatitis
22	976	93.4	191	2	AAR92940 Hepatitis
23	976	93.4	191	3	AAY94411 Human hep
24	976	93.4	249	8	AdJ10436 HCV Core
25	976	93.4	473	8	AdJ10438 HCV Core (

## ALIGNMENTS

## RESULT 1

AAR92987  
ID AAR92987 standard; protein; 191 AA.

XX AAR92987;

XX AC (first entry)

DT 02-OCT-1996 (first entry)

DE Hepatitis C virus isolate HK2 core protein.

XX HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;

KW hepatitis.

XX Hepatitis C virus.

XX WO9605315-A2.

PN 22-FEB-1996.

XX 15-AUG-1995; 95WO-US010398.

PF 15-AUG-1994; 94US-00290665.

XX (USSH ) US SEC DEPT HEALTH.

PA Bukh J, Miller RH, Purcell RH;

XX WPI; 1996-139709/14.

DR N-PSDB; AAT16661.

XX DNA and amino acid sequence of HCV envelope 1 and core proteins - used to

PT determine HCV genotype and as vaccines against HCV infection.

XX Claim 4; Page 223; 340pp; English.

CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV

CC isolates. Isolated cDNA sequences are used for the prodn. of primers

CC useful for detecting the presence of HCV in a sample, the primers are

CC also useful for HCV genotyping. Proteins encoded by the cDNAs can be used

CC in vaccines for immunising against HCV infection. The proteins may also

CC be used to detect antibodies against HCV in serum, saliva, lymphocytes or

CC other mononuclear cells. The antibodies may be used in the prevention of

XX HCV infection

XX Sequence 191 AA;

Query Match 100.0%; Score 1045; DB 2; Length 191;

Best Local Similarity 100.0%; Pred. No. 2.5e-90;

	Matches	191;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	MSTLPKFKQRTKRTNTRRP	TDVKFPFGGQ	I	VGGVYLL	PRRGPR	LGVRAT	KTRT	SRSQ	PRG 60
Db	1	MSTLPKFKQRTKRTNTRRP	TDVKFPFGGQ	I	VGGVYLL	PRRGPR	LGVRAT	KTRT	SRSQ	PRG 60
Qy	61	RRQIPKARQPGQHWAQPG	PWPPLY	NGEGCG	WGLLS	PRGSR	PHWG	PNDP	RRSR	NLG 120
Db	61	RRQIPKARQPGQHWAQPG	PWPPLY	NGEGCG	WGLLS	PRGSR	PHWG	PNDP	RRSR	NLG 120
Qy	121	KVDTLTTCGFADLMGYIP	VPVCGAP	LG	GVAAALAH	GVRA	IEDG	IN	VATGNL	PGCSFSIFLLA 180
Db	121	KVDTLTTCGFADLMGYIP	VPVCGAP	LG	GVAAALAH	GVRA	IEDG	IN	VATGNL	PGCSFSIFLLA 180
Qy	181	LLSCLTTP	PASA	191						
Db	181	LLSCLTTP	PASA	191						

Qy	61	RRQIPKARQPQGRHNAQPGYPWELYNMGEGCGAGWLLSPRGSRPHWGPNDRRPSRNIG	120
Dd	61	RRQIPKARQPEGRSWAQPGYPWELYNMGEGCGAGWLLSPRGSRPSWGNDRRPSRNIG	120
Qy	121	KVIDTLTCGFADLMGYIPVVCAPLGGVAALAHGVRAITEDGINYATGNLPGCSFSIFLLA	180
Dd	121	KVIDTLTCGFADLMGYIPIVGAPVGVARALAHGVRAVEDGINYATGNLPGCSFSIFLLA	180
Qy	181	LLSCLTTPASA 191	
Dd	181	LLSCLTVPASA 191	
RESULT 3			
AAR92974			
ID AAR92974		standard; protein; 191 AA.	
XX AC	AAR92974;		
XX DT	02-OCT-1996 (first entry)		
XX DE	Hepatitis C virus isolate Z1 core protein.		
XX KW	HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine; hepatitis.		
XX OS	Hepatitis C virus.		
XX PN	WO9605315-A2.		
XX PD	22-FEB-1996.		
XX PF	15-AUG-1995; 95WO-USO10398.		
XX PR	15-AUG-1994; 94US-00290665.		
XX PA	(USSH ) US SEC DEPT HEALTH.		
XX PI	Bukh J, Miller RH, Purcell RH;		
XX DR	WPI; 1996-139709/14.		
XX DR	N-PSDB; AAT16648.		
XX PT	DNA and amino acid sequence of HCV envelope 1 and core proteins - used to determine HCV genotype and as vaccines against HCV infection.		
XX PS	Claim 4; Page 212; 340pp; English.		
XX CC	AAR92936-R92987 are HCV core proteins derived from 52 different HCV isolates. Isolated cDNA sequences are used for the prodn. of primers useful for detecting the presence of HCV in a sample, the primers are also useful for HCV genotyping. Proteins encoded by the cDNAs can be used in vaccines for immunising against HCV infection. The proteins may also be used to detect antibodies against HCV in serum, saliva, lymphocytes or other mononuclear cells. The antibodies may be used in the prevention of HCV infection		
XX SO	Sequence 191 AA:		

	Query Match	95.3%	Score 996;	DB 2;	Length 191;
	Best Local Similarity	94.8%;	Pred. No. 1e-85;		
	Matches 181;	Conservative	4;	Mismatches 6;	Indels 0;
	Gaps 0;				
Qy	1	MSTLPQKRTKNTNRPRPTDVKPPGGQIVGGVYLLPRRGLVGRVATRTTSRSPRG	60		
Db	1	MSNTPKQKRTKNTNRPRMDVDFGGQIVGGVYLLPRRGLVGRVATRTTSRSPRG	60		

Db	1	MSTNPKPQRTKRNTRNRPMDVKPFPGGQIVGGVYLLPRRGPRLGVAARKTSERSQPRG	60
Qy	61	RRQTIKARQPCQGRHWAQGPYPWPLYGNEGCGWALLSPGRSGHMGPNDRPRRSRNLG	120
Db	61	RRQTIKARSGEGRSWAQPYPWPLYGNEGCGWALLSPGRSGRSGWPNDRPRRSRNLG	120
Qy	121	KVIDTLTCGFDLMDGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFIFLLA	180



XX 02-OCT-1996 (first entry)  
 XX Hepatitis C virus isolate DK13 core protein.  
 DE HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;  
 XX hepatitis.  
 XX Hepatitis C virus.  
 OS Hepatitis C virus.  
 XX WO9605315-A2.  
 XX 22-FEB-1996.  
 XX 15-AUG-1995; 95WO-US010398.  
 XX 15-AUG-1994; 94US-00290665.  
 XX (USSH ) US SEC DEPT HEALTH.  
 XX Bukh J, Miller RH, Purcell RH;  
 PI WPI; 1996-139709/14.  
 DR N-PSDB; AAT16652.  
 XX DNA and amino acid sequence of HCV envelope 1 and core proteins - used to  
 PT determine HCV genotype and as vaccines against HCV infection.  
 XX Claim 4; Page 215-216; 340pp; English.  
 XX AAR92936-R92987 are HCV core proteins derived from 52 different HCV  
 CC isolates. Isolated cDNA sequences are used for the prodn. of primers  
 CC useful for detecting the presence of HCV in a sample, the primers are  
 CC also useful for HCV genotyping. Proteins encoded by the cDNAs can be used  
 CC in vaccines for immunising against HCV infection. The proteins may also  
 CC be used to detect antibodies against HCV in serum, saliva, lymphocytes or  
 CC other mononuclear cells. The antibodies may be used in the prevention of  
 CC HCV infection  
 XX Sequence 191 AA;  
 SQ Query Match 93.8%; Score 980; DB 2; Length 191;  
 Best Local Similarity 93.7%; Pred. No. 3.3e-84;  
 Matches 179; Conservative 4; Mismatches 8; Indels 0; Gaps 0;  
 Qy 1 MSTILPKPQRTKNTNRPTDVKFPGGGQIVGGVYLLPRGPRGLGVATRAKTSERSQPRG 60  
 Db 1 MSTNPKPQRTKNTNRPRMDVKFPGGGQIVGGVYLLPRGPRGLGVATRAKTSERSQPRG 60  
 Qy 61 RRQIPKARQPOGRHWAQPGYPWPLYNEGCGWAGWLLSPRGSRPHWGPNDPRRSRNIG 120  
 Db 61 RRQIPKARQEGSWAOPGYPWPLYNEGCGWAGWLLSPRGSRPHWGPNDPRRSRNIG 120  
 Qy 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
 Db 121 KVIDTLTCGFADLMGYIPVVGAPVGGVARALAHGVRLLEDGVNATGNLPGCSFSIFLLA 180  
 Qy 181 LLSCLTTPASA 191  
 Db 181 LLSCLTTPASA 191  
 RESULT 7  
 AAR92976  
 ID AAR92976 standard; protein; 191 AA.  
 XX AAR92976;  
 XX 02-OCT-1996 (first entry)  
 XX Hepatitis C virus isolate Z6 core protein.  
 DE HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;  
 XX hepatitis.

KW hepatitis.  
 XX Hepatitis C virus.  
 OS WO9605315-A2.  
 XX 22-FEB-1996.  
 XX 15-AUG-1995; 95WO-US010398.  
 XX 15-AUG-1994; 94US-00290665.  
 XX (USSH ) US SEC DEPT HEALTH.  
 XX Bukh J, Miller RH, Purcell RH;  
 PI WPI; 1996-139709/14.  
 DR N-PSDB; AAT16650.  
 XX DNA and amino acid sequence of HCV envelope 1 and core proteins - used to  
 PT determine HCV genotype and as vaccines against HCV infection.  
 XX Claim 4; Page 214; 340pp; English.  
 XX AAR92936-R92987 are HCV core proteins derived from 52 different HCV  
 CC isolates. Isolated cDNA sequences are used for the prodn. of primers  
 CC useful for detecting the presence of HCV in a sample, the primers are  
 CC also useful for HCV genotyping. Proteins encoded by the cDNAs can be used  
 CC in vaccines for immunising against HCV infection. The proteins may also  
 CC be used to detect antibodies against HCV in serum, saliva, lymphocytes or  
 CC other mononuclear cells. The antibodies may be used in the prevention of  
 CC HCV infection  
 XX Sequence 191 AA;  
 SQ Query Match 93.6%; Score 978; DB 2; Length 191;  
 Best Local Similarity 93.2%; Pred. No. 5.1e-84;  
 Matches 178; Conservative 5; Mismatches 8; Indels 0; Gaps 0;  
 Qy 1 MSTILPKPQRTKNTNRPTDVKFPGGGQIVGGVYLLPRGPRGLGVATRAKTSERSQPRG 60  
 Db 1 MSTNPKPQRTKNTNRPRMDVKFPGGGQIVGGVYLLPRGPRGLGVATRAKTSERSQPRG 60  
 Qy 61 RRQIPKARQPOGRHWAQPGYPWPLYNEGCGWAGWLLSPRGSRPHWGPNDPRRSRNIG 120  
 Db 61 RRQIPKARSEGSWAOPGYPWPLYNEGCGWAGWLLSPRGSRPHWGPNDPRRSRNIG 120  
 Qy 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
 Db 121 KVIDTLTCGFADLMGYIPVVGAPVGGVARALAHGVRAVEDGINVATGNLPGCSFSIFLLA 180  
 Qy 181 LLSCLTTPASA 191  
 Db 181 LLSCLTTPASA 191  
 RESULT 8  
 AAR92977  
 ID AAR92977 standard; protein; 191 AA.  
 XX AAR92977;  
 XX 02-OCT-1996 (first entry)  
 XX Hepatitis C virus isolate Z7 core protein.  
 DE HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;  
 XX hepatitis.  
 OS Hepatitis C virus.  
 XX WO9605315-A2.  
 XX

PD 22-FEB-1996.  
 XX 15-AUG-1995; 95WO-US010398.  
 XX 15-AUG-1994; 94US-00290665.  
 XX (USSH ) US SEC DEPT HEALTH.  
 XX Bukh J, Miller RH, Purcell RH;  
 XX WPI; 1996-139709/14.  
 DR N-PSDB; AAT16651.  
 XX DNA and amino acid sequence of HCV envelope 1 and core proteins - used to  
 PT determine HCV genotype and as vaccines against HCV infection.  
 XX Claim 4; Page 214-215; 340pp; English.  
 XX AAR92936-R92987 are HCV core proteins derived from 52 different HCV  
 CC isolates. Isolated cDNA sequences are used for the prodn. of primers  
 CC useful for detecting the presence of HCV in a sample, the primers are  
 CC also useful for HCV genotyping. Proteins encoded by the cDNAs can be used  
 CC in vaccines for immunising against HCV infection. The proteins may also  
 CC be used to detect antibodies against HCV in serum, saliva, lymphocytes or  
 CC other mononuclear cells. The antibodies may be used in the prevention of  
 CC HCV infection  
 XX Sequence 191 AA;  
 SQ Query Match 93.5%; Score 977; DB 2; Length 191;  
 Best Local Similarity 93.2%; Pred. No. 6.3e-84;  
 Matches 178; Conservative 5; Mismatches 8; Indels 0; Gaps 0;  
 QY 1 MSTLPKQKTKRNTNRRPTDVKFGGGQIVGGVYLLPRGRPLGVRAATKTSERSQPRG 60  
 DB 1 MSTNPKPQKTKRNTNRRPDMVKFGGGQIVGGVYLLPRGRPLGVRAATKTSERSQPRG 60  
 QY 61 RQQIPKARQPOGRHWAQPGYPWPLYGNECGWAGWLLSPGRSRPHWGPNDRRRSRNLG 120  
 DB 61 RQQIPKARSRGSRWAQPGYPWPLYGNECGWAGWLLSPGRSRPHWGPNDRRRSRNLG 120  
 QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180  
 DB 121 KVIDTLTCGFADLMGYIPLVGAPVGVVARALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180  
 QY 181 LLSCLTTPASA 191  
 DB 181 LLSCLTVPASA 191  
 RESULT 9  
 AAR74048  
 ID AAR74048 standard; protein; 196 AA.  
 XX AAR74048;  
 XX 25-MAR-2003 (revised)  
 DT 26-NOV-1995 (first entry)  
 XX Synthetic HCV nucleocapsid protein.  
 XX Hepatitis C virus; vector; expression; prokaryotic cells; detection;  
 KW diagnosis.  
 XX Synthetic.  
 XX WO9511980-A2.  
 XX 04-MAY-1995.  
 XX 25-OCT-1994; 94WO-US012166.  
 XX 25-OCT-1993; 93US-00141917.  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX Khudyakov Y, Fields HA;  
 XX WPI; 1995-178872/23.  
 XX 25-OCT-1993; 93US-00141917.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX Khudyakov Y, Fields HA;  
 XX WPI; 1995-178872/23.  
 DR N-PSDB; AAQ92071.  
 XX Vector for expression of a synthetically produced protein coding DNA -  
 PT pref. encodes the hepatitis C virus nucleocapsid protein which can be  
 PT used in the detection of HCV antibodies.  
 XX Disclosure; Page 37; 44pp; English.  
 XX The sequence is that of a synthetic nucleotide sequence encoding the  
 CC hepatitis C virus nucleocapsid protein. The gene is positioned in a  
 CC vector for efficient expression in prokaryotic cells. The expressed  
 CC protein can be used in tests for the detection of antibodies specific for  
 CC the HCV C protein. See also AAR74047. (Updated on 25-MAR-2003 to correct  
 CC PN field.)  
 XX Sequence 196 AA;  
 SQ Query Match 93.5%; Score 977; DB 2; Length 196;  
 Best Local Similarity 92.1%; Pred. No. 6.5e-84;  
 Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;  
 QY 1 MSTLPKQKTKRNTNRRPTDVKFGGGQIVGGVYLLPRGRPLGVRAATKTSERSQPRG 60  
 DB 1 MSTIPKQKTKRNTNRRPQDVKFGGGQIVGGVYLLPRGRPLGVRAATKTSERSQPRG 60  
 QY 61 RQQIPKARQPOGRHWAQPGYPWPLYGNECGWAGWLLSPGRSRPHWGPNDRRRSRNLG 120  
 DB 61 RQQIPKVRPRPGRTWAQPGYPWPLYGNECGWAGWLLSPGRSRPHWGPNDRRRSRNLG 120  
 QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180  
 DB 121 KVIDTLTCGFADLMGYIPLVGAPLGGVAAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180  
 QY 181 LLSCLTTPASA 191  
 DB 181 LLSCLTVPASA 191  
 RESULT 10  
 AAR74047  
 ID AAR74047 standard; protein; 196 AA.  
 XX AAR74047;  
 XX 25-MAR-2003 (revised)  
 DT 26-NOV-1995 (first entry)  
 XX Synthetic HCV nucleocapsid protein.  
 XX Hepatitis C virus; vector; expression; prokaryotic cells; detection;  
 KW diagnosis.  
 XX Synthetic.  
 XX WO9511980-A2.  
 XX 04-MAY-1995.  
 XX 25-OCT-1994; 94WO-US012166.  
 XX 25-OCT-1993; 93US-00141917.  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX Khudyakov Y, Fields HA;  
 XX WPI; 1995-178872/23.

DR N-PSDB; AAQ92070.

XX Vector for expression of a synthetically produced protein coding DNA -

PT pref. encodes the hepatitis C virus nucleocapsid protein which can be

PT used in the detection of HCV antibodies.

XX

XX Disclosure; Page 35; 44pp; English.

XX

CC The sequence is that of a synthetic nucleotide sequence encoding the

CC hepatitis C virus nucleocapsid protein. The gene is positioned in a

CC vector for efficient expression in prokaryotic cells. The expressed

CC protein can be used in tests for the detection of antibodies specific for

CC the HCV C protein. See also AAR74048. (Updated on 25-MAR-2003 to correct

CC PN field.)

XX

SQ Sequence 196 AA;

Query Match 93.5%; Score 977; DB 2; Length 196;

Best Local Similarity 92.1%; Pred. No. 6.5e-84;

Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

Qy 1 MSTLPKQRTKNTNRPTDVKFPGGQIVGGVYLLPRGPRGLGVRATKTSRSQPRG 60

Db 1 MSTLPKQRTKNTNRPTDVKFPGGQIVGGVYLLPRGPRGLGVRATKTSRSQPRG 60

Qy 61 RRQPIPKAROPQGRHWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPHWGPNDRSRNLG 120

Db 61 RRQPIPKAROPQGRHWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPHWGPNDRSRNLG 120

Qy 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPCCSFSIFLLA 180

Db 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPCCSFSIFLLA 180

Qy 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

RESULT 11

AAR96546

ID AAR96546 standard; peptide; 319 AA.

XX

AC AAR96546;

XX

DT 10-MAR-1997 (first entry)

XX

DE Hepatitis C virus types 9a(7a) isolates FRI amino acids 1-317.

XX

KW Hepatitis C virus; subtype; polymerase chain reaction; amplification;

KW PCR; primer; probe; antibody; infection.

XX

OS Hepatitis C virus.

XX

Key Location/Qualifiers

FT Misc-difference 128

FT /label= Phe, Ser, Tyr, Cys

FT Misc-difference 129

FT /label= Gly

FT /note= "amino acid in this position is designated X in

FT the specification, but codon usage shows that the only

FT possible amino acid at this pos. is Gly"

FT Misc-difference 308

FT /label= Ile, Met

FT Misc-difference 310

FT /label= Ser, Pro, Thr, Ala

XX

PN W09613590-A2.

XX

PD 09-MAY-1996.

XX

PF 23-OCT-1995; 95WO-EP004155.

XX

PR 21-OCT-1994; 94EP-00870166.

PR 28-JUN-1995; 95EP-00870076.

XX (INNO-) INNOGENETICS NV.

PA

XX Maertens G, Stuyver L;

PI

XX WPI; 1996-251460/25.

DR

XX N-PSDB; AAT27957.

DR

XX

XX Hepatitis C virus polynucleic acid unique to unidentified sub-type -

PT used to develop probes and primers for new subtypes and vaccines to

PT prevent and treat infection.

XX

XX Claim 25; Fig 3; 150pp; English.

PS

XX

CC The sequences AAR96526-R96578 represent novel sequences isolated from

CC hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f, 4a-

CC j, 5a and 6a. They esp. from the novel subtypes id-f, 2e-i, 2k, 2l, 3g,

CC 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5'

CC untranslated region (UR), the Core/E1, NS4 or NS5B regions of the genome.

CC This sequence represents amino acids 1-317 from the HCV types 9a and 7a

CC isolates FRI. The new HCV types were isolated from patients with chronic

CC HCV from the Benelux countries, France, Cameroon and Vietnam, because of

CC their aberrant reactivities. The RNA was extracted, cDNA synthesised and

CC PCR amplified, cloned and genotyped. The 5'UR, Core/E1 and NS5B regions

CC were sequenced either directly or partially and used to classify the new

CC viruses into (sub)types based on comparison with known sequences. The

CC sequences were also used to generate the peptides AAR96424-R96524. The

CC sequences can also be used to synthesise probes and primers for the

CC detection of HCV in a sample. The polypeptides can be used to detect anti

CC -HCV antibodies, for HCV typing or to prevent HCV infections

XX

SQ Sequence 319 AA;

Query Match 93.5%; Score 977; DB 2; Length 319;

Best Local Similarity 93.7%; Pred. No. 1.1e-83;

Matches 179; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

Qy 1 MSTLPKQRTKNTNRPTDVKFPGGQIVGGVYLLPRGPRGLGVRATKTSRSQPRG 60

Db 1 MSTLPKQRTKNTNRPTDVKFPGGQIVGGVYLLPRGPRGLGVRATKTSRSQPRG 60

Qy 61 RRQPIPKAROPQGRHWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPHWGPNDRSRNLG 120

Db 61 RRQPIPKAROPQGRHWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPHWGPNDRSRNLG 120

Qy 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPCCSFSIFLLA 180

Db 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPCCSFSIFLLA 180

Qy 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

RESULT 12

AAR22137

ID AAR22137 standard; protein; 326 AA.

XX

AC AAR22137;

XX

DT 27-AUG-2003 (revised)

DT 25-MAR-2003 (revised)

DT 07-JUL-1992 (first entry)

XX

DE HCV-Hc59 capsid and envelope proteins.

XX

KW Hepatitis C virus; non-A non-B virus; HCV-Hc59; antigen; vaccine; assay;

KW detection.

XX

OS Non-A.

OS non-B hepatitis virus.

XX

```

FH Key      Location/Qualifiers
FT Region   1..326
FT           /label= NANBV_structural_proteins
FT Region   1..120
FT           /label= capsid
FT Peptide  1..74
FT           /label= pref._capsid_antigen
FT Peptide  1..20
FT           /label= pref._capsid_antigen
FT Peptide  2..40
FT           /label= pref._capsid_antigen
FT Misc-difference 3
FT           /label= Ile
FT           /note= "or Asn according to Seq No 46 (AAR22154), see CC"
FT Peptide  21..40
FT           /label= pref._capsid_antigen
FT Peptide  69..120
FT           /label= pref._capsid_antigen
FT Region   121..326
FT           /label= envelope_protein
FT Peptide  121..176
FT           /label= pref._envelope_antigen
FT Misc-difference 321
FT           /label= Asp
FT           /note= "or Asn according to Seq No 1 (AAR22154), see CC"
FT W09203458-A.
FT
FT 05-MAR-1992.
FT
FT 23-AUG-1991; 91WO-US006037.
FT
FT 25-AUG-1990; 90US-00573643.
FT 21-NOV-1990; 90US-00616369.
FT 21-AUG-1991; 91US-00748564.
FT
FT (NYBL-) NEW YORK BLOO DCENT.
FT (PHAR-) PHARMA.
FT
FT Zebedee S, Inchauspe G, Nasofe MS, Prince AM;
FT
FT WPI; 1992-096821/12.
FT N-PSDB; AAQ22838.
FT
FT Deoxyribonucleic acid sequence encoding non-A, non-B hepatitis virus -
FT obtd. Hutch C59 subgroup encoding polypeptide(s), useful as vaccines, and
FT immuno reactive ABS for diagnosis of virus.
FT
FT Disclosure; Page 131-133; 225pp; English.
FT
FT One Hutch strain (HCV-H) of NANBV, designated the Hutch c59 isolate (HCV-
FT HC59) was propagated through passage in animals and the entire viral
FT genome was cloned and sequenced (see AAQ22871). The sequence represented
FT here comprises two amino acids, indicated in the features, which differ
FT from the sequence of AAR22154. The proteins and peptides (see features)
FT and antibodies against them are useful for the prepn. of vaccines and
FT inoculums against NANBV and in immunological assays for detection of
FT viral infection. (Updated on 25-MAR-2003 to correct PR field.) (Updated
FT on 25-MAR-2003 to correct PA field.) (Updated on 27-AUG-2003 to correct
FT OS field.)
FT
FT Sequence 326 AA;
FT
FT Query Match      93.5%; Score 977; DB 2; Length 326;
FT Best Local Similarity 92.1%; Pred. No. 1.1e-83;
FT Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;
FT
FT 1 MSTLPKPKRKNTNRRPTDVKFPGGGQIVGGVLLPRRGLGVRATRKTSERSQPRG 60
FT
FT 1 MSTLPKPKRKNTNRRPTDVKFPGGGQIVGGVLLPRRGLGVRATRKTSERSQPRG 60
FT
FT 61 RQPIPKARQPOGRHWAQPGYPWPVLYGNEGCWAGWLLSPRGRPHWGPNDRRRSRNLG 120
FT
FT 121 KVIDTLTCGFADLMGYIPVUGAPLGGVAAALAHGVRAEDGYNATGNLPGCSFSIFLLA 180
FT
FT 121 KVIDTLTCGFADLMGYIPVUGAPLGGVAAALAHGVRAEDGYNATGNLPGCSFSIFLLA 180
FT
FT 181 LLSCLTTPASA 191
FT
FT 181 LLSCLTTPASA 191

```

```

Db      61 RQPIPKARPEGRRTWAQPGYPWPVLYGNEGCWAGWLLSPRGRSRPSWGPTDPRRRSRNLG 120
Qy      121 KVIDTLTCGFADLMGYIPVUGAPLGGVAAALAHGVRAEDGYNATGNLPGCSFSIFLLA 180
Db      121 KVIDTLTCGFADLMGYIPVUGAPLGGVAAALAHGVRAEDGYNATGNLPGCSFSIFLLA 180
Qy      181 LLSCLTTPASA 191
Db      181 LLSCLTTPASA 191

RESULT 13
AAR38278
ID AAR38278 standard; protein; 733 AA.
XX AAR38278;
XX
XX 21-OCT-1993 (first entry)
XX
XX NANB hepatitis virus HC-OM gene polypeptide P-733-1.
XX
XX Non-A, non-B; virus; polymerase chain reaction; detection; sensitive;
XX specific; HCV; NANBH.
XX
XX Non-A.
XX non-B hepatitis virus.
XX
XX JP05091884-A.
XX
XX 16-APR-1993.
XX
XX 10-APR-1991; 91JP-00196175.
XX
XX 12-JUN-1990; 90JP-00153401.
XX 08-NOV-1990; 90JP-00304405.
XX
XX (NAKA/) NAKAMURA T.
XX
XX WPI; 1993-199637/25.
XX N-PSDB; AAQ43888.
XX
XX Antigen related to non-A and non-B hepatitis virus - comprises non-
XX translation region comprising 340 - 341 mols. of nucleotides, non-
XX translation region comprising 1885 - 2551 mols. of nucleotides including
XX region 1,149 and, etc.
XX
XX Claim 14; Page 32-35; 73pp; Japanese.
XX
XX The sequence is that of NANB hepatitis virus HC-OM gene polypeptide P-733
XX -1. It may be used in a system for detecting NANB hepatitis. This
XX method is highly specific and sensitive, and can detect NANB hepatitis
XX virus which could not be detected by conventional methods
XX
XX Sequence 733 AA;
XX
XX Query Match      93.5%; Score 977; DB 2; Length 733;
XX Best Local Similarity 92.1%; Pred. No. 2.8e-83;
XX Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;
Qy      1 MSTLPKPKRKNTNRRPTDVKFPGGGQIVGGVLLPRRGLGVRATRKTSERSQPRG 60
Db      1 MSTLPKPKRKNTNRRPTDVKFPGGGQIVGGVLLPRRGLGVRATRKTSERSQPRG 60
Qy      61 RQPIPKARQPOGRHWAQPGYPWPVLYGNEGCWAGWLLSPRGRPHWGPNDRRRSRNLG 120
Db      61 RQPIPKVRPEGRRTWAQPGYPWPVLYGNEGCWAGWLLSPRGRSRPSWGPTDPRRRSRNLG 120
Qy      121 KVIDTLTCGFADLMGYIPVUGAPLGGVAAALAHGVRAEDGYNATGNLPGCSFSIFLLA 180
Db      121 KVIDTLTCGFADLMGYIPVUGAPLGGVAAALAHGVRAEDGYNATGNLPGCSFSIFLLA 180
Qy      181 LLSCLTTPASA 191

```





PT and blood products.  
XX  
PS Disclosure; Fig 1; 51pp; English.  
XX  
CC AAR70230 is the composite hepatitis C virus (HC-J1/CDC/CHI) protein from  
CC which the synthetic HCV antigens described in AAR70210-R70229 were  
CC derived. These synthetic antigens can be used to screen blood, or blood  
CC products for the presence of HCV, they can also be used in various specific  
CC assays for the detection of HCV antibodies, and antigens, or as  
CC immunogens. (Updated on 25-MAR-2003 to correct FN field.) (Updated on 25-  
CC MAR-2003 to correct PF field.)  
XX  
SQ Sequence 2894 AA;

Query Match 93.5%; Score 977; DB 2; Length 2894;  
Best Local Similarity 92.1%; Pred. No. 1.3e-82;  
Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;  
QY 1 MSTLPKPKQKTKRNTNRRPTDVKFGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60  
DB :|||||  
1 MSTLPKPKQKTKRNTNRRPQDVKFGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60  
QY 61 RRQIPKARQPGQRHWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPHWGPNDRSRNLG 120  
DB :|||:|  
61 RRQIPKVRPRPGRTWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPTDPRRSRNLG 120  
QY 121 KVIDTLTCGFADLMGYIPVVGAPLGVAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180  
DB :|||||  
121 KVIDTLTCGFADLMGYIPVVGAPLGVAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180  
QY 181 LLSCLTTPASA 191  
DB :|||||  
181 LLSCLTVPASA 191

Search completed: October 30, 2004, 01:42:44  
Job time : 157 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2004, 01:35:06 ; Search time 39 Seconds  
(without alignments)  
471.215 Million cell updates/sec

Title: US-09-084-691B-206

Perfect score: 1045

Sequence: 1 MSTLPKQKTKRNNRRPT.....CSFSIFLLALLSCLTTPASA 191

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_79: \*  
1: PIR1: \*  
2: PIR2: \*  
3: PIR3: \*  
4: PIR4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	977	93.5	3011	1	S40770	genome polyprotein
2	976	93.4	3011	1	GNWVCH	genome polyprotein
3	971	92.9	441	2	S12707	genome polyprotein
4	969	92.7	782	2	S18031	genome polyprotein
5	969	92.7	787	2	PN0677	hypothetical prote
6	968	92.6	3011	1	GNWVC3	genome polyprotein
7	965	92.3	3010	1	GNWVCJ	genome polyprotein
8	965	92.3	3010	1	GNWVC	genome polyprotein
9	962	92.1	550	2	JH0711	genome polyprotein
10	961	92.0	640	2	JQ1584	genome polyprotein
11	959	91.8	189	2	S32740	polyprotein - hepa
12	958	91.7	782	2	S19876	genome polyprotein
13	957	91.6	513	2	PC1284	genome polyprotein
14	957	91.6	782	2	S19875	genome polyprotein
15	957	91.6	876	2	PC2219	polypeptide - hepa
16	956	91.5	3010	1	GNWVTW	genome polyprotein
17	956	91.5	3010	1	S18030	genome polyprotein
18	952	91.1	520	2	JQ1925	polyprotein - hepa
19	951	91.0	411	2	PC2060	genome polyprotein
20	948	90.7	3010	1	A45573	genome polyprotein
21	947	90.6	369	2	S21471	genome polyprotein
22	947	90.6	523	2	JQ1926	genome polyprotein
23	947	90.6	782	2	S18032	genome polyprotein
24	946	90.5	492	2	S41288	genome polyprotein
25	944	90.3	411	2	PC2061	genome polyprotein
26	941	90.0	874	2	JQ0883	genome polyprotein
27	940	90.0	513	2	A44150	structural protein
28	939	89.9	874	2	JQ0881	genome polyprotein
29	939	89.9	3033	1	JQ1303	genome polyprotein

#### ALIGNMENTS

##### RESULT 1

S40770

genome polyprotein - hepatitis C virus

N:Contains: capsid protein C; envelope protein M; hepatitis protein NS5 (nonstructu

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 09-Jul-2004

C:Accession: S40770; PC1285

R:Okamoto, H.

submitted to the EMBL Data Library, March 1992

A:Reference number: S40770

A:Accession: S40770

A:Molecule type: genomic RNA

A:Residues: 1-3011 <OKA>

A:Cross-references: UNIPROT:Q03463; EMBL:D10749; NID:g221586; PIDN:BAA01582.1; PID:g2215

R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda,

Jun. J. Exp. Med. 60, 167-177, 1990

A:Title: The 5'-terminal sequence of the hepatitis C virus genome.

A:Reference number: PC1284; MCID:91013116; PMID:2170712

A:Accession: PC1285

A:Molecule type: genomic RNA

A:Residues: 1-513 <OK2>

A:Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512

A:Experimental source: isolate HC-J1

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serin

F:2-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <MEE>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepatitis virus #status predicted <NS3>

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1618-1862/Product: nonstructural protein NS4a #status predicted <NS4>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>

F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match

Best Local Similarity 92.1%; Pred. No. 2,9e-74;

Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKQKTKRNNRRPTDVKFPGGQIVGGVYLLPRRGLGVRAATKTSERSQPRG 60

Db 1 MSTLPKQKTKRNNRRPTDVKFPGGQIVGGVYLLPRRGLGVRAATKTSERSQPRG 60

QY 61 RRQIPKARQPCRWAOQGYWPVLYGNECCWAGWLLSPRGRPHWGNDRRRSRNLG 120

Db 61 RRQIPKARQPCRWAOQGYWPVLYGNECCWAGWLLSPRGRPHWGNDRRRSRNLG 120

QY 121 KVIDTLTCGFADLMGVIPVWGAPLGVAAALAHGVRAIEDGINATGNLPGCSFSIFLLA 180



QY 181 LLSCLTTPASA 191  
Db 181 LLSCLTTPASA 191

RESULT 5  
PN0677  
hypotheoretical protein 787 - hepatitis C virus (fragment)  
C:Species: hepatitis C virus  
C:Date: 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change 09-Jul-2004  
C:Accession: PN0677  
R:Cho, S.H.; Yoon, J.I.; Chang, J.E.; Ahn, B.M.; Lee, C.H.; Lee, Y.I.  
Biochem. Biophys. Res. Commun. 196; 780-788, 1993  
A:Title: Genomic typing of hepatitis C viruses from Korean patients: Implications of gen  
A:Reference number: PN0677; PMID:94059104; PMID:8240354  
A:Accession: PN0677  
A:Molecule type: mRNA  
A:Residues: 1-787 <CHO>  
A:Cross-references: UNIPROT:O08244; GB:L20498; NID:gl381031; PIDN:AA02608.1; PID:gl3810  
C:Superfamily: hepatitis C virus genome polypeptide  
C:Keywords: glycoprotein; nonstructural protein  
F:196,209,234,250,305,325,421,427,452,536,544,560,580,627,649/Binding site: carbohydrate  
Query Match 92.7%; Score 969; DB 2; Length 787;  
Best Local Similarity 92.7%; Pred. No. 3.8e-74;  
Matches 177; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKNTNRPTDVKFGGQIVGGVYLLPRGRPLGVRATKTSERSQPRG 60  
Db 1 MSTNPKPKRKNTNRPPQDVKFGGQIVGGVYLLPRGRPLGVRATKTSERSQPRG 60

QY 61 RROPTPKARQPGRWAOQGYWPVLYGEGCCWAGWLLSPGSRPHWGPNDRSRNLG 120  
Db 61 RROPTPKARQPGRWAOQGYWPVLYGEGCCWAGWLLSPGSRPHWGPNDRSRNLG 120

QY 121 KVIDLTGCFADLMGYIPVWGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
Db 121 KVIDLTGCFADLMGYIPVWGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191  
Db 181 LLSCLTTPASA 191

RESULT 6  
GNWVCJ  
genome polypeptide - hepatitis C virus (strain HCV-1)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructu  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 09-Jul-2004  
C:Accession: A39166; PQ0404  
R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coi  
Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991  
A:Title: Genetic organization and diversity of the hepatitis C virus.  
A:Reference number: A39166; PMID:91172826; PMID:1848704  
A:Accession: A39166  
A:Molecule type: mRNA  
A:Residues: 1-3011 <CHO>  
A:Cross-references: UNIPROT:P26664; GB:M62321; NID:g329873; PIDN:AAA45676.1; PID:g329874  
R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.I  
J. Gen. Virol. 73, 1131-1141, 1992  
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e  
A:Reference number: PQ0393; PMID:92268871; PMID:1316939  
A:Accession: PQ0403  
A:Molecule type: genomic RNA  
A:Residues: 1577-1633 <CHA>  
A:Cross-references: DDBJ:D10128  
A:Experimental source: isolates E-b16  
A:Accession: PQ0404  
A:Status: preliminary  
A:Molecule type: genomic RNA

A:Residues: 1577-1633 <CH2>  
A:Experimental source: isolates E-b17  
C:Superfamily: hepatitis C virus genome polypeptide; glycoprotein; hydrolase; nonstructura  
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura  
F:1-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEB>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepatitis C virus genome polypeptide <NS3>  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4>  
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 92.6%; Score 968; DB 1; Length 3011;  
Best Local Similarity 91.6%; Pred. No. 1.6e-73;  
Matches 175; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKNTNRPTDVKFGGQIVGGVYLLPRGRPLGVRATKTSERSQPRG 60  
Db 1 MSTNPKPKRKNTNRPPQDVKFGGQIVGGVYLLPRGRPLGVRATKTSERSQPRG 60

QY 61 RROPTPKARQPGRWAOQGYWPVLYGEGCCWAGWLLSPGSRPHWGPNDRSRNLG 120  
Db 61 RROPTPKARQPGRWAOQGYWPVLYGEGCCWAGWLLSPGSRPHWGPNDRSRNLG 120

QY 121 KVIDLTGCFADLMGYIPVWGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
Db 121 KVIDLTGCFADLMGYIPVWGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191  
Db 181 LLSCLTTPASA 191

RESULT 7  
GNWVCJ  
genome polypeptide - hepatitis C virus (strain J)  
N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructura  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 09-Jul-2004  
C:Accession: A39253; PS0086  
R:Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimot  
Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990  
A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients  
A:Reference number: A39253; PMID:91088550; PMID:2175903  
A:Accession: A39253  
A:Molecule type: genomic RNA  
A:Residues: 1-3010 <KAT>  
A:Cross-references: UNIPROT:P26662; GB:D90208; NID:g221610; PIDN:BAAL4233.1; PID:g221611  
R:Kato, N.; Ohkoshi, S.; Shimotohno, K.  
Proc. Jpn. Acad. 65B, 219-223, 1989  
A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence vari  
A:Reference number: PS0085  
A:Accession: PS0086  
A:Molecule type: genomic RNA  
A:Residues: 2650-2707 <KAT>  
A:Experimental source: Japanese isolate  
C:Comment: The cleavage sites of this polypeptide have not been determined.  
C:Superfamily: hepatitis C virus genome polypeptide  
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serin  
F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEB>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepatitis C virus genome polypeptide <NS3>  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F;1312-1317/Region: nucleotide-binding motif B  
F;1316-1319/Region: DEXH motif

F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>  
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>  
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F;196,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2

Query Match 92.3%; Score 965; DB 1; Length 3010;  
Best Local Similarity 92.1%; Pred. No. 2.9e-73;  
Matches 176; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

Qy 1 MSTLPKQRTKRNTRPTDVKPPGGQIVGGVYLLPRRGLGVRAATKTSRSQPRG 60

Db 1 MSTNPKQRTKRNTRPPQDKPPGGQIVGGVYLLPRRGLGVRAATKTSRSQPRG 60

Qy 61 RRQIPKARPOQRHWAQPGYPWPLYNCGWAGWLLSPRGSPHGWPNDRPRRSNLG 120

Db 61 RRQIPKARPEGRWTWAPQGPWPLYNCGWAGWLLSPRGSPHGWPNDRPRRSNLG 120

Qy 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPCSCSIFLLA 180

Db 121 KVIDTLTCGFADLMGYIPLVGAPLGGVAAALAHGVRAIEDGINVATGNLPCSCSIFLLA 180

Qy 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

#### RESULT 8

GNWYTC

genome polyprotein - hepatitis C virus  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein NS5  
Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C;Species: hepatitis C virus

C;Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 09-Jul-2004

C;Accession: A38465

R;Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.;

J. Virol. 65, 1105-1113, 1991

A;Title: Structure and organization of the hepatitis C virus genome isolated from human

A;Reference number: A38465; MUID:91140698; PMID:1847440

A;Accession: A38465

A;Molecule type: genomic RNA

A;Residues: 1-3010 <TA>

A;Cross-references: UNIPROT:P26663; EMBL:M58335; NID:G329770; PIDN:AAA72945.1; PID:G3297

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: ARF; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural

F;2-115/Product: capsid protein C #status predicted <CPC>

F;116-191/Product: envelope protein M #status predicted <BPM>

F;192-389/Product: major envelope protein E #status predicted <MEE>

F;730-729/Product: nonstructural protein NS1 #status predicted <NS1>

F;1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>

F;1230-1237/Product: hepatitis C virus genome polyprotein NS3

F;1312-1317/Region: nucleotide-binding motif A (P-loop)

F;1316-1319/Region: DEXH motif

F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>

F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>

F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

F;196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 92.3%; Score 965; DB 1; Length 3010;

Best Local Similarity 92.1%; Pred. No. 2.9e-73;

Matches 176; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

Qy 1 MSTLPKQRTKRNTRPTDVKPPGGQIVGGVYLLPRRGLGVRAATKTSRSQPRG 60

Db 1 MSTNPKQRTKRNTRPPQDKPPGGQIVGGVYLLPRRGLGVRAATKTSRSQPRG 60

Qy 61 RRQIPKARPOQRHWAQPGYPWPLYNCGWAGWLLSPRGSPHGWPNDRPRRSNLG 120

Db 61 RRQIPKARPEGRWTWAPQGPWPLYNCGWAGWLLSPRGSPHGWPNDRPRRSNLG 120

Qy 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPCSCSIFLLA 180

Db 121 KVIDTLTCGFADLMGYIPLVGAPLGGVAAALAHGVRAIEDGINVATGNLPCSCSIFLLA 180

Qy 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

#### RESULT 9

JH0711

genome polyprotein - hepatitis C virus (strain PRC1) (fragments)

N:Contains: envelope protein E1; envelope protein E2; nonstructural protein NS1; nonstru

C;Species: hepatitis C virus

C;Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 09-Jul-2004

C;Accession: JH0711

R;Liu, K.; Hu, Z.; Li, H.; Prince, A.M.; Inchauspe, G.

Gene 114, 245-250, 1992

A;Title: Genomic typing of hepatitis C viruses present in China.

A;Reference number: JH0711; MUID:92290283; PMID:1318245

A;Accession: JH0711

A;Molecule type: genomic RNA

A;Residues: 1-550 <LIU>

A;Cross-references: UNIPROT:Q7LZV4; GB:M74888; GB:M74889

A;Note: the nucleotide sequence is not complete

A;Note: translation of the nucleotide sequence is not complete

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: envelope protein; glycoprotein; nonstructural protein; nucleocapsid; polypro

F;1-190/Product: nucleocapsid protein C #status predicted <CPC>

F;191-380/Product: envelope protein E1 #status predicted <EPI>

F;381-514/Product: envelope protein E2 or nonstructural protein NS1 (fragment) #status

F;515-550/Product: nonstructural protein NS5 (fragment) #status predicted <NS5>

F;196,233,250,305,416,422,429,447/Binding site: carbohydrate (Asn) #status pr

Query Match 92.1%; Score 962; DB 2; Length 550;

Best Local Similarity 91.6%; Pred. No. 1.1e-73;

Matches 175; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

Qy 1 MSTLPKQRTKRNTRPTDVKPPGGQIVGGVYLLPRRGLGVRAATKTSRSQPRG 60

Db 1 MSTNPKQRTKRNTRPPQDKPPGGQIVGGVYLLPRRGLGVRAATKTSRSQPRG 60

Qy 61 RRQIPKARPOQRHWAQPGYPWPLYNCGWAGWLLSPRGSPHGWPNDRPRRSNLG 120

Db 61 RRQIPKARPEGRWTWAPQGPWPLYNCGWAGWLLSPRGSPHGWPNDRPRRSNLG 120

Qy 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPCSCSIFLLA 180

Db 121 KVIDTLTCGFADLMGYIPLVGAPLGGVAAALAHGVRAIEDGINVATGNLPCSCSIFLLA 180

Qy 181 LLSCLTTPASA 191

Db 181 LLSCLTTPASA 191

#### RESULT 10

JQ1584

genome polyprotein - hepatitis C virus (strain U.K.) (fragment)

N:Contains: core protein C; envelope protein E1; envelope protein E2; nonstructural pro

C;Species: hepatitis C virus

C;Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 09-Jul-2004

C;Accession: JQ1584

R;Kumar, U.; Cheng, D.; Thomas, H.; Monjardino, J.

J. Gen. Virol. 73, 1521-1525, 1992

A;Title: Cloning and sequencing of the structural region and expression of putative cor

A;Reference number: JQ1584; MUID:92300349; PMID:1318944

A;Accession: JQ1584

A;Molecule type: genomic RNA

A;Residues: 1-640 <KUM>

A;Cross-references: UNIPROT:Q68966; GB:X84079; NID:G643119; PIDN:CAA58888.1; PID:G64312

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: core protein; envelope protein; glycoprotein; nonstructural protein; polypro

F;1-191/Product: core protein C #status predicted <CPC>

F;192-389/Product: envelope protein E1 #status predicted <EE1>

F:390-640/Product: envelope protein E2 and nonstructural protein NS1 #status predicted &  
F:196,209,234,305,417,430,448,476,540,556,576,623/Binding site: carbohydrate (asn) (cov)

Query Match 92.0%; Score 961; DB 2; Length 640;  
Best Local Similarity 91.1%; Pred. No. 1.5e-73;  
Matches 174; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

QY 1 MSTLPKPKQKTKRNTNRRPTDVKFPGGGQIVGGVLLPRRGLGVRATRKTSERSQPRG 60  
DB 1 MSTNPKPQKTKRNTNRRPDQVKFPGGGQIVGGVLLPRRGLGVRATRKTSERSQPRG 60  
QY 61 RROPIPKARQPOGRHWAQPGYPWPVLYGNEGCGWAGWLLSPRGRPHWPNDRSRNLG 120  
DB 61 RROPIPKARPEGRRTWAQPGYPWPVLYGNEGCGWAGWLLSPRGRSPSWGPTDPRSRNLG 120  
QY 121 KVIDLTTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180  
DB 121 KVIDLTTCGFADLMGYIPLVGLGGAARALAHGVRAIEDGVNYATGNLPGCSFSIFLLA 180  
QY 181 LLSCLTTPASA 191  
DB 181 LLSCLTTPASA 191

## RESULT 11

S32740

polyprotein - hepatitis C virus (isolate Russian) (fragment)

N:Contains: capsid protein C; envelope protein M

C:Species: hepatitis C virus

C>Date: 19-Mar-1997 #sequence\_revision 05-Dec-1998 #text\_change 09-Jul-2004

C:Accession: S32740

R:Vassilev, V.B.; Viazov, S.O.; Kotova, E.Y.; Nosikov, V.V.

submitted to the EMBL Data Library, April 1993

A:Description: Evidence of new HCV variant of European isolate in Russia.

A:Reference number: S32740

A:Accession: S32740

A:Molecule type: genomic RNA

A:Residues: 1-189 <VAS>

A:Cross-references: UNIPROT:Q68873; EMBL:X71407

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; envelope protein; polyprotein

F:1-115/Product: capsid protein C #status predicted <CPC>

F:116-189/Product: envelope protein M #status predicted <EPM>

Query Match 91.8%; Score 959; DB 2; Length 189;

Best Local Similarity 92.1%; Pred. No. 6.9e-74;  
Matches 174; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 1 MSTLPKPKQKTKRNTNRRPTDVKFPGGGQIVGGVLLPRRGLGVRATRKTSERSQPRG 60  
DB 1 MSTNPKPQKTKRNTNRRPDQVKFPGGGQIVGGVLLPRRGLGVRATRKTSERSQPRG 60  
QY 61 RROPIPKARQPOGRHWAQPGYPWPVLYGNEGCGWAGWLLSPRGRPHWPNDRSRNLG 120  
DB 61 RROPIPKARPEGRRTWAQPGYPWPVLYGNEGCGWAGWLLSPRGRSPSWGPTDPRSRNLG 120  
QY 121 KVIDLTTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180  
DB 121 KVIDLTTCGFADLMGYIPLVGLGGAARALAHGVRAIEDGVNYATGNLPGCSFSIFLLA 180  
QY 181 LLSCLTTPA 189  
DB 181 LLSCLTTPA 189

## RESULT 12

S19876

genome polyprotein - hepatitis C virus (isolate JK5) (fragment)

N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein

C:Species: hepatitis C virus

A:Variety: isolate JK5

C>Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 09-Jul-2004

C:Accession: S19876

R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.

submitted to the EMBL Data Library, September 1991

A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso

A:Reference number: S18029

A:Accession: S19876

A:Molecule type: genomic RNA

A:Residues: 1-782 <HON>

A:Cross-references: UNIPROT:Q68953; EMBL:X61595; NID:g59486; PIDN:CAA43792.1; PID:g59487

A:Experimental source: isolate JK5

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural

F:1-191/Product: core protein #status predicted <MAT1>

F:192-383/Product: envelope protein 1 #status predicted <MAT2>

F:384-733/Product: NS1/E2 protein #status predicted <MAT3>

F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 91.7%; Score 958; DB 2; Length 782;

Best Local Similarity 91.6%; Pred. No. 3.2e-73;  
Matches 175; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 1 MSTLPKPKQKTKRNTNRRPTDVKFPGGGQIVGGVLLPRRGLGVRATRKTSERSQPRG 60  
DB 1 MSTNPKPQKTKRNTNRRPDQVKFPGGGQIVGGVLLPRRGLGVRATRKTSERSQPRG 60  
QY 61 RROPIPKARQPOGRHWAQPGYPWPVLYGNEGCGWAGWLLSPRGRPHWPNDRSRNLG 120  
DB 61 RROPIPKARPEGRRTWAQPGYPWPVLYGNEGCGWAGWLLSPRGRSPSWGPTDPRSRNLG 120  
QY 121 KVIDLTTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180  
DB 121 KVIDLTTCGFADLMGYIPLVGLGGAARALAHGVRAIEDGVNYATGNLPGCSFSIFLLA 180  
QY 181 LLSCLTTPASA 191  
DB 181 LLSCLTTPASA 191

## RESULT 13

PC1284

genome polyprotein - hepatitis C virus (isolate HC-J4) (fragment)

C:Species: hepatitis C virus

C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 09-Jul-2004

C:Accession: PC1284

R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda,

Upn. J. Exp. Med. 60, 167-177, 1990

A:Title: The 5'-terminal sequence of the hepatitis C virus genome.

A:Reference number: PC1284; MUID:91013116; PMID:2170712

A:Accession: PC1284

A:Molecule type: genomic RNA

A:Residues: 1-513 <OK>

A:Cross-references: UNIPROT:Q81221; GB:D00832; NID:g221513; PIDN:BAA00706.1; PID:g221514

C:Superfamily: hepatitis C virus genome polyprotein

Query Match

Best Local Similarity 91.6%; Score 957; DB 2; Length 513;

Matches 175; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

QY 1 MSTLPKPKQKTKRNTNRRPTDVKFPGGGQIVGGVLLPRRGLGVRATRKTSERSQPRG 60  
DB 1 MSTNPKPQKTKRNTNRRPDQVKFPGGGQIVGGVLLPRRGLGVRATRKTSERSQPRG 60  
QY 61 RROPIPKARQPOGRHWAQPGYPWPVLYGNEGCGWAGWLLSPRGRPHWPNDRSRNLG 120  
DB 61 RROPIPKARPEGRRTWAQPGYPWPVLYGNEGCGWAGWLLSPRGRSPSWGPTDPRSRNLG 120  
QY 121 KVIDLTTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180  
DB 121 KVIDLTTCGFADLMGYIPLVGLGGAARALAHGVRAIEDGVNYATGNLPGCSFSIFLLA 180  
QY 181 LLSCLTTPASA 191  
DB 181 LLSCLTTPASA 191





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2004, 01:34:01 ; Search time 195 Seconds

(without alignments)  
563.572 Million cell updates/sec

Title: US-09-084-691B-206

Perfect score: 1045

Sequence: 1 MSTLPKPKQKRNTRPT.....CSFSIFLLALLSCLTPASA 191

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 02:\*

1: uniprot\_sprot:\*

2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1045	100.0	191	2	Q68115
2	1035	99.0	191	2	Q68411
3	1025	98.1	414	2	P89956
4	1024	98.0	414	2	P89958
5	1024	98.0	414	2	Q81329
6	1024	98.0	3016	2	O92531
7	1023	97.9	3018	2	O39927
8	1020	97.6	414	2	P89963
9	1016	97.2	414	2	P89957
10	1016	97.2	414	2	P89960
11	1014	97.0	414	2	P89959
12	1011	96.7	414	2	Q81270
13	1009	96.6	319	2	Q68703
14	1009	96.6	414	2	P89962
15	1007	96.4	191	2	O93057
16	1007	96.4	414	2	Q81269
17	1003	96.0	414	2	Q81268
18	1003	96.0	414	2	Q81314
19	1001	95.8	191	2	O70635
20	1000	95.7	414	2	Q81271
21	999	95.6	414	2	P89955
22	997	95.4	414	2	P89954
23	997	95.4	415	2	P89961
24	997	95.4	415	2	P89964
25	997	95.4	415	2	P89965
26	997	95.4	3015	2	O92532
27	996	95.3	191	2	Q68153
28	995	95.2	415	2	Q81272
29	995	95.2	415	2	Q81315
30	990	94.7	415	2	P87751
31	990	94.7	3013	2	O92530

```

32 988 94.5 414 2 Q81267
33 988 94.5 514 2 Q6XJ47
34 988 94.5 514 2 AAP69953
35 988 94.5 3019 2 O92529
36 987 94.4 191 2 P87841
37 987 94.4 415 2 Q81274
38 985 94.3 416 2 O39647
39 984 94.2 319 2 Q68698
40 984 94.2 415 2 Q81273
41 982 94.0 191 2 Q68152
42 982 94.0 191 2 Q68157
43 982 94.0 319 2 Q68705
44 981 93.9 191 2 Q68116
45 981 93.9 191 2 Q68576

```

#### ALIGNMENTS

```

RESULT 1
Q68115 PRELIMINARY; PRT; 191 AA.
AC Q68115;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Polypeptide (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]_TaxID=11103;
RP SEQUENCE FROM N.A.
RA MEDLINE=94336721; PubMed=8058787;
RX Bukh J., Purcell R.H., Miller R.H.;
RT "Sequence analysis of the core gene of 14 hepatitis C virus
   genotypes.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:8239-8243(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92279243; PubMed=1317578;
RA Bukh J., Purcell R.H., Miller R.H.;
RT "Sequence analysis of the 5' noncoding region of hepatitis C virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:4942-4946(1992).
RN [3]
RP SEQUENCE FROM N.A.
RA Bukh J., Purcell R.H., Miller R.H.;
RT "At least 12 genotypes of hepatitis C virus predicted by sequence
   analysis of the putative E1 gene of isolates collected worldwide.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:8234-8238(1993).
DR EMBL; U0198; AAA21037.1; -.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002522; HCV capsid.
DR Pfam; PF01543; HCV_capsid; 1.
DR Polyprotein.
FT CHAIN 1 >191 core protein.
FT NON_TER 191 191
SQ SEQUENCE 191 AA; 20725 MW; 7A7994DF2F0909C6 CRC64;

```

Query Match 100.0%; Score 1045; DB 2; Length 191;

Best Local Similarity 100.0%; Pred. No. 7e-80;

Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSTLPKPKQKRNTRPTDVKFPGGQIVGGVYLLPRGPRGLGVRATRKTSERSQPRG 60

DB 1 MSTLPKPKQKRNTRPTDVKFPGGQIVGGVYLLPRGPRGLGVRATRKTSERSQPRG 60

QY 61 RROPIPKARQPGRHWAQPGVWPVLYGNCGWAGWLLSPRGSRPHWGPNDRPRSRNLG 120

Db 61 RRQPIKARQPGQHRHWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPHWGPNDPRRSRNLG 120  
QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
Db 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
QY 181 LLSCLTTPASA 191  
Db 181 LLSCLTTPASA 191

RESULT 2  
Q68411 PRELIMINARY; PRT; 191 AA.  
AC Q68411;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Core protein (Fragment).  
OS Hepatitis C virus type 6.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=42182;  
[1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=97052554; PubMed=8897188;  
RA Bernier L., Williams B., Delage G., Murphy D.G.;  
RT "Identification of numerous hepatitis C virus genotypes in Montreal,  
Canada."  
RL J. Clin. Microbiol. 34:2815-2818 (1996).  
DR EMBL; U33435; AAB40038.1; -.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR002522; HCV\_capsid.  
DR InterPro; IPR002521; HCV\_core.  
DR Pfam; PF01543; HCV\_core; 1.  
DR Pfam; PF01542; HCV\_core; 1.  
FT NON TER 191\_191  
SQ SEQUENCE 191 AA; 20789 MW; F292AB64B56DE30A CRC64;

Query Match 99.0%; Score 1035; DB 2; Length 191;  
Best Local Similarity 99.0%; Pred. No. 4.8e-79;  
Matches 189; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MSTLPKQPKTKNTNRPTDVKPPGGQIVGGVYLLPRGPRLGVRATKTSERSQPRG 60  
Db 1 MSTLPKQPKTKNTNRPTDVKPPGGQIVGGVYLLPRGPRLGVRATKTSERSQPRG 60  
QY 61 RRQPIKARQPGQHRHWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPHWGPNDPRRSRNLG 120  
Db 61 RRQPIKARQPGQHRHWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPHWGPNDPRRSRNLG 120  
QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
Db 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
QY 181 LLSCLTTPASA 191  
Db 181 LLSCLTTPASA 191

RESULT 3  
P89956 PRELIMINARY; PRT; 414 AA.  
AC P89956;  
DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Core, env and part of E2/NS1 (Fragment).  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
[1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=95062197; PubMed=7972001;  
RA Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,  
Iizuka H., Mishihiro S., Miyakawa Y., Mayumi M.;  
RT "Hepatitis C virus variants from Vietnam are classifiable into the  
seventh, eighth, and ninth major genetic groups."  
Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026 (1994).  
[2]  
RN

RESULT 4  
P89958 PRELIMINARY; PRT; 414 AA.  
AC P89958;  
DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Core, env and part of E2/NS1 (Fragment).  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
[1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=95062197; PubMed=7972001;  
RA Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,  
Iizuka H., Mishihiro S., Miyakawa Y., Mayumi M.;  
RT "Hepatitis C virus variants from Vietnam are classifiable into the  
seventh, eighth, and ninth major genetic groups."  
Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026 (1994).  
[2]  
RN

RN SEQUENCE FROM N.A.  
RP MEDLINE=95062197; PubMed=7972001;  
RX Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,  
Iizuka H., Mishihiro S., Miyakawa Y., Mayumi M.;  
RT "Hepatitis C virus variants from Vietnam are classifiable into the  
seventh, eighth, and ninth major genetic groups."  
Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026 (1994).  
[2]  
RN SEQUENCE FROM N.A.  
RP Okamoto H.;  
RA Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.  
RL EMBL; D88469; BAA13618.1; -.  
DR PIR; P00804; P00804.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; F:structural molecule activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR002522; HCV\_capsid.  
DR InterPro; IPR002521; HCV\_core.  
DR InterPro; IPR002519; HCV\_core.  
DR InterPro; IPR002531; HCV\_NS1.  
DR Pfam; PF01543; HCV\_capsid; 1.  
DR Pfam; PF01542; HCV\_core; 1.  
DR Pfam; PF01539; HCV\_core; 1.  
DR Pfam; PF01560; HCV\_NS1; 1.  
DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
KW Polyprotein; Transmembrane.  
FT NON TER 414\_414  
SQ SEQUENCE 414 AA; 44528 MW; 7A3BE1710311C017 CRC64;

Query Match 98.1%; Score 1025; DB 2; Length 414;  
Best Local Similarity 97.9%; Pred. No. 7.3e-78;  
Matches 187; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 1 MSTLPKQPKTKNTNRPTDVKPPGGQIVGGVYLLPRGPRLGVRATKTSERSQPRG 60  
Db 1 MSTLPKQPKTKNTNRPTDVKPPGGQIVGGVYLLPRGPRLGVRATKTSERSQPRG 60  
QY 61 RRQPIKARQPGQHRHWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPHWGPNDPRRSRNLG 120  
Db 61 RRQPIKARQPGQHRHWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPHWGPNDPRRSRNLG 120  
QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
Db 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
QY 181 LLSCLTTPASA 191  
Db 181 LLSCLTTPASA 191

RESULT 4  
P89958 PRELIMINARY; PRT; 414 AA.  
AC P89958;  
DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Core, env and part of E2/NS1 (Fragment).  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
[1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=95062197; PubMed=7972001;  
RA Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,  
Iizuka H., Mishihiro S., Miyakawa Y., Mayumi M.;  
RT "Hepatitis C virus variants from Vietnam are classifiable into the  
seventh, eighth, and ninth major genetic groups."  
Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026 (1994).  
[2]  
RN

InterPro: IPR002521; HCV core.  
InterPro: IPR002519; HCV env.  
InterPro: IPR002531; HCV NS1.  
Pfam: PF01543; HCV capsid; 1.  
Pfam: PF01542; HCV core; 1.  
Pfam: PF01539; HCV env; 1.  
Pfam: PF01560; HCV NS1; 1.  
Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
Polyprotein; Transmembrane.  
NON\_TER 414 414  
SEQUENCE 414 AA; 44937 MW; 65355640863B3DCF CRC64;

Query Match 98.0%; Score 1024; DB 2; Length 414;  
Best Local Similarity 98.4%; Pred. No. 8.8e-78;  
Matches 188; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSTLPKPKQKTKRNTNRRPTDVKPPGGQIVGGVYLLPRRGLRVATRTKTSRSQPRG 60  
DB 1 MSTLPKPKQKTKRNTNRRPMDVKFPGGQIVGGVYLLPRRGLRVATRTKTSRSQPRG 60  
QY 61 RRQPIPKARQPGGRHWAQGYWPPLYGNEGCGWALLSPRGRPHWGPNDPRRSRNLG 120  
DB 61 RRQPIPKARQSGRHWAGPYWPPLYGNEGCGWALLSPRGRSPNPGNDPRRSRNLG 120  
QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
DB 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
QY 181 LLSCLTTPASA 191  
DB 181 LLSCLTTPASA 191

RESULT 6  
Q92531 PRELIMINARY; PRT; 3016 AA.  
ID Q92531  
AC Q92531;  
DT 01-NOV-1998 (TREMELrel. 08, Created)  
DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)  
DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)  
DE Polyprotein.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98378034; PubMed=9714232;  
RA Tokita H., Okamoto H., Iizuka H., Kishimoto J., Tsuda F., Miyakawa Y.,  
RA Mayumi M.;  
RT "The entire nucleotide sequences of three hepatitis C virus isolates  
RT in genetic groups 7-9 and comparison with those in the other eight  
RT genetic groups."  
RL J. Gen. Virol. 79:1847-1857(1998).  
DR EMBL; D84264; BAA32666.1; -;  
DR FIP; PQ0804; PQ0804.  
DR HSP; P27958; LAIV.  
DR MEROPS; S29.001; -;  
DR MEROPS; U39.001; -;  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006350; P:transcription; IEA.  
DR GO; GO:0019079; P:viral genome replication; IEA.  
DR GO; GO:0019087; P:viral transformation; IEA.  
DR InterPro: IPR00345; CytC heme BS.

InterPro: IPR002521; HCV core.  
InterPro: IPR002519; HCV env.  
InterPro: IPR002531; HCV NS1.  
Pfam: PF01543; HCV capsid; 1.  
Pfam: PF01542; HCV core; 1.  
Pfam: PF01539; HCV env; 1.  
Pfam: PF01560; HCV NS1; 1.  
Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
Polyprotein; Transmembrane.  
NON\_TER 414 414  
SEQUENCE 414 AA; 44973 MW; F3F3CF154372P4FF CRC64;

Query Match 98.0%; Score 1024; DB 2; Length 414;  
Best Local Similarity 97.4%; Pred. No. 8.8e-78;  
Matches 186; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSTLPKPKQKTKRNTNRRPTDVKPPGGQIVGGVYLLPRRGLRVATRTKTSRSQPRG 60  
DB 1 MSTLPKPKQKTKRNTNRRPMDVKFPGGQIVGGVYLLPRRGLRVATRTKTSRSQPRG 60  
QY 61 RRQPIPKARQPGGRHWAQGYWPPLYGNEGCGWALLSPRGRPHWGPNDPRRSRNLG 120  
DB 61 RRQPIPKARQPTGRHWAQGYWPPLYGNEGCGWALLSPRGRPHWGPNDPRRSRNLG 120  
QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
DB 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
QY 181 LLSCLTTPASA 191  
DB 181 LLSCLTTPASA 191

RESULT 5  
Q81329 PRELIMINARY; PRT; 414 AA.  
ID Q81329  
AC Q81329;  
DT 01-NOV-1996 (TREMELrel. 01, Created)  
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Core, env and part of E2/NS1 (Fragment).  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95062197; PubMed=7972001;  
RA Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,  
RA Iizuka H., Mishihiro S., Miyakawa Y., Mayumi M.;  
RT "Hepatitis C virus variants from Vietnam are classifiable into the  
RT seventh, eighth, and ninth major genetic groups."  
RN Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Okamoto H.;  
RA Submitted (AUG-1996) to the EMBL/GenBank/DBSJ databases.  
DR EMBL; D88468; BAA13617.1; -;  
DR FIP; PQ0804; PQ0804.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro: IPR002522; HCV capsid.

InterPro: IPR002521; HCV core.  
InterPro: IPR002519; HCV env.  
InterPro: IPR002531; HCV NS1.  
Pfam: PF01543; HCV capsid; 1.  
Pfam: PF01542; HCV core; 1.  
Pfam: PF01539; HCV env; 1.  
Pfam: PF01560; HCV NS1; 1.  
Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
Polyprotein; Transmembrane.  
NON\_TER 414 414  
SEQUENCE 414 AA; 44973 MW; F3F3CF154372P4FF CRC64;

Query Match 98.0%; Score 1024; DB 2; Length 414;  
Best Local Similarity 97.4%; Pred. No. 8.8e-78;  
Matches 186; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSTLPKPKQKTKRNTNRRPTDVKPPGGQIVGGVYLLPRRGLRVATRTKTSRSQPRG 60  
DB 1 MSTLPKPKQKTKRNTNRRPMDVKFPGGQIVGGVYLLPRRGLRVATRTKTSRSQPRG 60  
QY 61 RRQPIPKARQPGGRHWAQGYWPPLYGNEGCGWALLSPRGRPHWGPNDPRRSRNLG 120  
DB 61 RRQPIPKARQPTGRHWAQGYWPPLYGNEGCGWALLSPRGRPHWGPNDPRRSRNLG 120  
QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
DB 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
QY 181 LLSCLTTPASA 191  
DB 181 LLSCLTTPASA 191

RESULT 5  
Q81329 PRELIMINARY; PRT; 414 AA.  
ID Q81329  
AC Q81329;  
DT 01-NOV-1996 (TREMELrel. 01, Created)  
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Core, env and part of E2/NS1 (Fragment).  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95062197; PubMed=7972001;  
RA Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,  
RA Iizuka H., Mishihiro S., Miyakawa Y., Mayumi M.;  
RT "Hepatitis C virus variants from Vietnam are classifiable into the  
RT seventh, eighth, and ninth major genetic groups."  
RN Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Okamoto H.;  
RA Submitted (AUG-1996) to the EMBL/GenBank/DBSJ databases.  
DR EMBL; D88468; BAA13617.1; -;  
DR FIP; PQ0804; PQ0804.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro: IPR002522; HCV capsid.

DR InterPro: IPR001410; DEAD.  
DR InterPro: IPR002522; HCV capsid.  
DR InterPro: IPR002521; HCV\_core.  
DR InterPro: IPR002519; HCV env.  
DR InterPro: IPR002531; HCV NS1.  
DR InterPro: IPR000745; HCV NS4a.  
DR InterPro: IPR001490; HCV NS4b.  
DR InterPro: IPR002868; HCV NS5a.  
DR InterPro: IPR002166; HCV NS5a.  
DR InterPro: IPR004109; Peptidase S29.  
DR InterPro: IPR009003; Pept Ser Cys.  
DR InterPro: IPR002518; Pept\_U39 HCV NS2.  
DR InterPro: IPR007095; RNA pol DS PS.  
DR InterPro: IPR007094; RNA pol PSVir.  
DR Pfam: PF01543; HCV capsid; 1.  
DR Pfam: PF01542; HCV core; 1.  
DR Pfam: PF01539; HCV env; 1.  
DR Pfam: PF01538; HCV NS1; 1.  
DR Pfam: PF01537; HCV NS2; 1.  
DR Pfam: PF02907; HCV NS3; 1.  
DR Pfam: PF01006; HCV NS4a; 1.  
DR Pfam: PF01001; HCV NS4b; 1.  
DR Pfam: PF01506; HCV NS5a; 1.  
DR Pfam: PF00998; Viral RdRP; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.  
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
KW Polyprotein; Transmembrane.  
SQ SEQUENCE 3016 AA; 328032 MW; 4ESCFF96258BCE3B CRC64;

Query Match 98.0%; Score 1024; DB 2; Length 3016;  
Best Local Similarity 98.4%; Pred. No. 6.6e-77;  
Matches 188; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MSTLPKPKQRTKNTNRPTDVKFPGGQIVGGVYLLPRRGLGVRAATKTSRSQPRG 60  
Db 1 MSTLPKPKQRTKNTNRPTDVKFPGGQIVGGVYLLPRRGLGVRAATKTSRSQPRG 60

Qy 61 RROPIPKARQPGQRHWAQPGVWPPLYGNEGCGWAGWLLSPRGRPHWGNPDPRRSNRLG 120  
Db 61 RROPIPKARQPGQRHWAQPGVWPPLYGNEGCGWAGWLLSPRGRPHWGNPDPRRSNRLG 120

Qy 121 KVIDTLTCGFADLMGYIPVVGAPLGVAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
Db 121 KVIDTLTCGFADLMGYIPVVGAPLGVAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180

Qy 181 LLSCLTTPASA 191  
Db 181 LLSCLTTPASA 191

Query Match 97.9%; Score 1023; DB 2; Length 3018;  
Best Local Similarity 97.9%; Pred. No. 8e-77;  
Matches 187; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MSTLPKPKQRTKNTNRPTDVKFPGGQIVGGVYLLPRRGLGVRAATKTSRSQPRG 60  
Db 1 MSTLPKPKQRTKNTNRPTDVKFPGGQIVGGVYLLPRRGLGVRAATKTSRSQPRG 60

Qy 61 RROPIPKARQPGQRHWAQPGVWPPLYGNEGCGWAGWLLSPRGRPHWGNPDPRRSNRLG 120  
Db 61 RROPIPKARQPGQRHWAQPGVWPPLYGNEGCGWAGWLLSPRGRPHWGNPDPRRSNRLG 120

Qy 121 KVIDTLTCGFADLMGYIPVVGAPLGVAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
Db 121 KVIDTLTCGFADLMGYIPVVGAPLGVAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180

Qy 181 LLSCLTTPASA 191  
Db 181 LLSCLTTPASA 191

RESULT 8  
P89963 PRELIMINARY; PRT; 414 AA.  
ID P89963  
AC P89963;  
DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)

DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)  
DE Core, env and part of E2/NS1 (Fragment).  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95062197; PubMed=7972001;  
RA Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,  
Iizuka H., Mishihiro S., Miyakawa Y., Mayumi M.;  
RA "Hepatitis C virus variants from Vietnam are classifiable into the  
RT seventh, eighth, and ninth major genetic groups.";  
RT Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026(1994).  
RL [2]  
RN SEQUENCE FROM N.A.  
RP Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.  
RX EMBL; D88470; BAA13625.1; -;  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026(1994).  
RL [2]  
RN SEQUENCE FROM N.A.  
RP Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.  
RX EMBL; D88470; BAA13625.1; -;  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR009100; AcylCoA dehyd\_NM.  
DR InterPro; IPR002522; HCV capsid.  
DR InterPro; IPR002521; HCV core.  
DR InterPro; IPR002519; HCV env.  
DR InterPro; IPR002531; HCV NS1.  
DR Pfam; PF01543; HCV capsid; 1.  
DR Pfam; PF01542; HCV core; 1.  
DR Pfam; PF01539; HCV env; 1.  
DR Pfam; PF01560; HCV NS1; 1.  
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
KW Polyprotein; Transmembrane.  
FT NON\_TER 414  
SQ SEQUENCE 414 AA; 44804 MW; 397D0DC5C4BDA20D CRC64;  
  
Query Match 97.6%; Score 1020; DB 2; Length 414;  
Best Local Similarity 96.9%; Pred. No. 1.9e-77;  
Matches 185; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 MSTLPKQKTKRNTNRRPTDVKFGGGQIVGGVLLPRRGLGVRAATKTSERSQPRG 60  
DB 1 MSTLPKQKTKRNTNRRPTDVKFGGGQIVGGVLLPRRGLGVRAATKTSERSQPRG 60  
  
QY 61 RRQIPKARQPGQRHWAQGYGYPWLYGNEGCGWAGLLSPRGRPHWGPNDRRRSRNLG 120  
DB 61 RRQIPKARQPGQRHWAQGYGYPWLYGNEGCGWAGLLSPRGRPHWGPNDRRRSRNLG 120  
  
QY 121 KVDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINATGNLPGCSFSIFLLA 180  
DB 121 KVDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINATGNLPGCSFSIFLLA 180  
  
QY 181 LLSCLTTPASA 191  
DB 181 LLSCLTTPASA 191  
  
RESULT 9  
P89957  
ID P89957 PRELIMINARY; PRT; 414 AA.  
AC P89957;  
DT 01-MAY-1997 (TReMBLrel. 03, Created)  
DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)  
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)  
DE Core, env and part of E2/NS1 (Fragment).  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95062197; PubMed=7972001;  
  
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)  
DE Core, env and part of E2/NS1 (Fragment).  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95062197; PubMed=7972001;  
Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,  
Iizuka H., Mishihiro S., Miyakawa Y., Mayumi M.;  
RA "Hepatitis C virus variants from Vietnam are classifiable into the  
RT seventh, eighth, and ninth major genetic groups.";  
RT Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026(1994).  
RL [2]  
RN SEQUENCE FROM N.A.  
RP Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.  
RX EMBL; D88473; BAA13622.1; -;

RA Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,  
Iizuka H., Mishihiro S., Miyakawa Y., Mayumi M.;  
RT "Hepatitis C virus variants from Vietnam are classifiable into the  
seventh, eighth, and ninth major genetic groups.";  
Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026(1994).  
RL [2]  
RN SEQUENCE FROM N.A.  
RP Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.  
RX EMBL; D88470; BAA13619.1; -;  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR002522; HCV capsid.  
DR InterPro; IPR002521; HCV core.  
DR InterPro; IPR002519; HCV env.  
DR InterPro; IPR002531; HCV NS1.  
DR Pfam; PF01543; HCV capsid; 1.  
DR Pfam; PF01542; HCV core; 1.  
DR Pfam; PF01539; HCV env; 1.  
DR Pfam; PF01560; HCV NS1; 1.  
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
KW Polyprotein; Transmembrane.  
FT NON\_TER 414  
SQ SEQUENCE 414 AA; 44850 MW; 1754DD583F52414 CRC64;  
  
Query Match 97.2%; Score 1016; DB 2; Length 414;  
Best Local Similarity 96.9%; Pred. No. 4.1e-77;  
Matches 185; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 MSTLPKQKTKRNTNRRPTDVKFGGGQIVGGVLLPRRGLGVRAATKTSERSQPRG 60  
DB 1 MSTLPKQKTKRNTNRRPTDVKFGGGQIVGGVLLPRRGLGVRAATKTSERSQPRG 60  
  
QY 61 RRQIPKARQPGQRHWAQGYGYPWLYGNEGCGWAGLLSPRGRPHWGPNDRRRSRNLG 120  
DB 61 RRQIPKARQPGQRHWAQGYGYPWLYGNEGCGWAGLLSPRGRPHWGPNDRRRSRNLG 120  
  
QY 121 KVDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINATGNLPGCSFSIFLLA 180  
DB 121 KVDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINATGNLPGCSFSIFLLA 180  
  
QY 181 LLSCLTTPASA 191  
DB 181 LLSCLTTPASA 191  
  
RESULT 10  
P89960  
ID P89960 PRELIMINARY; PRT; 414 AA.  
AC P89960;  
DT 01-MAY-1997 (TReMBLrel. 03, Created)  
DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)  
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
DE Core, env and part of E2/NS1 (Fragment).  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95062197; PubMed=7972001;  
Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,  
Iizuka H., Mishihiro S., Miyakawa Y., Mayumi M.;  
RA "Hepatitis C virus variants from Vietnam are classifiable into the  
RT seventh, eighth, and ninth major genetic groups.";  
Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026(1994).  
RL [2]  
RN SEQUENCE FROM N.A.  
RP Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.  
RX EMBL; D88473; BAA13622.1; -;



```
Db 1 MSTLPKPKQKTKNTNRRPDVKFGGQIVGGVILLPRRGLGVRAATKTSERSQPRG 60
QY 61 RRQPIPKARQPOGRHWAQPGYWPVLYGNECGWAGWLLSPRGRPHWPNDRRRSRNLG 120
Db 61 RRQPIPKARHQTGRSQAQPGYWPVLYGNECGWAGWLLSPRGRPHWPNDRRRSRNLG 120
QY 121 KVIDLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180
Db 121 KVIDLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

RESULT 13
Q68703
ID Q68703 PRELIMINARY; PRT; 319 AA.
AC Q68703;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE Core protein (Fragment).
OS Hepatitis C virus type 6a.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus; Hepatitis C virus type 6.
OX NCBI_TaxID=31655;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96118171; PubMed=8578855;
RA Stuyver L., Wyseur A., van Arnhem W., Lunel F., Laurent-Puig P.,
RA Pawlotsky J.-M., Kleter B., Bassit L., Nkengasong J., van Doorn L.-J.,
RA Maertens G.;
RT "Hepatitis C virus genotyping by means of 5'-UR/core line probe assays
RT and molecular analysis of untypeable samples.";
RL Virus Res. 38:137-157(1995).
DR EMBL; L38339; AAC42193.1; -
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV env.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01539; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
KW Transmembrane.
FT CHAIN 1 191 core protein.
FT CHAIN 192 >319 E1 protein.
FT NON_TER 319 319
SQ SEQUENCE 319 AA; 34545 MW; DC1D0EA52ED64A8F CRC64;

Query Match 96.6%; Score 1009; DB 2; Length 319;
Best Local Similarity 96.3%; Pred. No. 1.2e-76;
Matches 184; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MSTLPKPKQKTKNTNRRPDVKFGGQIVGGVILLPRRGLGVRAATKTSERSQPRG 60
Db 1 MSTLPKPKQKTKNTNRRPDVKFGGQIVGGVILLPRRGLGVRAATKTSERSQPRG 60
QY 61 RRQPIPKARQPOGRHWAQPGYWPVLYGNECGWAGWLLSPRGRPHWPNDRRRSRNLG 120
Db 61 RRQPIPKARQPOGRHWAQPGYWPVLYGNECGWAGWLLSPRGRPHWPNDRRRSRNLG 120
QY 121 KVIDLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180
Db 121 KVIDLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

RESULT 14
Q68703
ID Q68703 PRELIMINARY; PRT; 414 AA.
AC Q68703;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DE Core, env and part of E2/NS1 (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95062197; PubMed=7972001;
RA Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,
RA Iizuka H., Mishiro S., Miyakawa Y., Mayumi M.;
RT "Hepatitis C virus variants from Vietnam are classifiable into the
RT seventh, eighth, and ninth major genetic groups.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026(1994).
DR EMBL; L38339; AAC42193.1; -
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 414 414
SQ SEQUENCE 414 AA; 44751 MW; 4BBA61438D11E3C6 CRC64;

Query Match 96.6%; Score 1009; DB 2; Length 414;
Best Local Similarity 96.3%; Pred. No. 1.6e-76;
Matches 184; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSTLPKPKQKTKNTNRRPDVKFGGQIVGGVILLPRRGLGVRAATKTSERSQPRG 60
Db 1 MSTLPKPKQKTKNTNRRPDVKFGGQIVGGVILLPRRGLGVRAATKTSERSQPRG 60
QY 61 RRQPIPKARQPOGRHWAQPGYWPVLYGNECGWAGWLLSPRGRPHWPNDRRRSRNLG 120
Db 61 RRQPIPKARQPOGRHWAQPGYWPVLYGNECGWAGWLLSPRGRPHWPNDRRRSRNLG 120
QY 121 KVIDLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180
Db 121 KVIDLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

RESULT 15
Q93057
ID Q93057 PRELIMINARY; PRT; 191 AA.
AC Q93057; Q68747;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
```

```
Db 181 LLSCLTTPASA 191
RESULT 14
P89962
ID P89962 PRELIMINARY; PRT; 414 AA.
AC P89962;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DE Core, env and part of E2/NS1 (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95062197; PubMed=7972001;
RA Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,
RA Iizuka H., Mishiro S., Miyakawa Y., Mayumi M.;
RT "Hepatitis C virus variants from Vietnam are classifiable into the
RT seventh, eighth, and ninth major genetic groups.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026(1994).
DR EMBL; L38339; AAC42193.1; -
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 414 414
SQ SEQUENCE 414 AA; 44751 MW; 4BBA61438D11E3C6 CRC64;

Query Match 96.6%; Score 1009; DB 2; Length 414;
Best Local Similarity 96.3%; Pred. No. 1.6e-76;
Matches 184; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSTLPKPKQKTKNTNRRPDVKFGGQIVGGVILLPRRGLGVRAATKTSERSQPRG 60
Db 1 MSTLPKPKQKTKNTNRRPDVKFGGQIVGGVILLPRRGLGVRAATKTSERSQPRG 60
QY 61 RRQPIPKARQPOGRHWAQPGYWPVLYGNECGWAGWLLSPRGRPHWPNDRRRSRNLG 120
Db 61 RRQPIPKARQPOGRHWAQPGYWPVLYGNECGWAGWLLSPRGRPHWPNDRRRSRNLG 120
QY 121 KVIDLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180
Db 121 KVIDLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

RESULT 15
Q93057
ID Q93057 PRELIMINARY; PRT; 191 AA.
AC Q93057; Q68747;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
```





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2004, 01:36:27 ; Search time 39 Seconds  
(without alignments)  
324.788 Million cell updates/sec

Title: US-09-084-691B-206  
Perfect score: 1045  
Sequence: 1 MSTLPKPKYKRNTRRPT.....CSFSIFLLALLSCLTTPASA 191

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A-COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B-COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A-COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B-COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PTUS-COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1045	100.0	191	2	US-08-290-665A-206
2	1045	100.0	191	5	PCT-US95-10398-206
3	1045	100.0	319	4	US-08-635-886C-228
4	1045	100.0	319	4	US-08-974-690C-228
5	996	95.3	191	2	US-08-290-665A-191
6	996	95.3	191	5	PCT-US95-10398-191
7	982	94.0	191	2	US-08-290-665A-192
8	982	94.0	191	2	US-08-290-665A-193
9	982	94.0	191	5	PCT-US95-10398-192
10	982	94.0	191	5	PCT-US95-10398-193
11	981	93.9	191	2	US-08-290-665A-172
12	981	93.9	191	5	PCT-US95-10398-172
13	980	93.8	191	2	US-08-290-665A-197
14	980	93.8	191	5	PCT-US95-10398-197
15	980	93.8	319	4	US-08-635-886C-217
16	980	93.8	319	4	US-08-974-690C-217
17	978	93.6	191	2	US-08-290-665A-195
18	978	93.6	191	5	PCT-US95-10398-195
19	977	93.5	191	2	US-08-290-665A-196
20	977	93.5	191	5	PCT-US95-10398-196
21	977	93.5	319	3	US-08-836-075A-42
22	977	93.5	319	4	US-08-635-886C-229
23	977	93.5	319	4	US-08-974-690C-229
24	977	93.5	450	4	US-08-635-886C-181
25	977	93.5	450	4	US-08-974-690C-181
26	977	93.5	2894	2	US-08-466-975A-23
27	977	93.5	2894	2	US-08-391-671A-23

28 977 93.5 2894 3 US-08-467-902A-23 Sequence 23, Appl  
29 977 93.5 2894 3 US-09-275-265-23 Sequence 23, Appl  
30 977 93.5 2894 4 US-09-941-611-23 Sequence 23, Appl  
31 976 93.4 191 2 US-08-290-665A-156 Sequence 156, Appl  
32 976 93.4 191 2 US-08-290-665A-157 Sequence 157, Appl  
33 976 93.4 191 2 US-08-290-665A-158 Sequence 158, Appl  
34 976 93.4 191 2 US-08-290-665A-159 Sequence 159, Appl  
35 976 93.4 191 2 US-08-290-665A-160 Sequence 160, Appl  
36 976 93.4 191 3 US-08-380-160-3 Sequence 3, Appl  
37 976 93.4 191 5 PCT-US95-10398-156 Sequence 156, Appl  
38 976 93.4 191 5 PCT-US95-10398-157 Sequence 157, Appl  
39 976 93.4 191 5 PCT-US95-10398-158 Sequence 158, Appl  
40 976 93.4 191 5 PCT-US95-10398-159 Sequence 159, Appl  
41 976 93.4 191 5 PCT-US95-10398-160 Sequence 160, Appl  
42 976 93.4 450 4 US-08-635-886C-179 Sequence 179, Appl  
43 976 93.4 450 4 US-08-635-886C-180 Sequence 180, Appl  
44 976 93.4 450 4 US-08-974-690C-179 Sequence 179, Appl  
45 976 93.4 450 4 US-08-974-690C-180 Sequence 180, Appl

## ALIGNMENTS

RESULT 1  
US-08-290-665A-206  
; Sequence 206, Application US/08290665A  
; Patent No. 5882852  
; GENERAL INFORMATION:  
; APPLICANT: BUKH, J., MILLER, R.H. AND  
; APPLICANT: PURCELL, R.H.  
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND  
; TITLE OF INVENTION: CORE GENES OF HEPATITIS C VIRUS  
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE  
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES  
; NUMBER OF SEQUENCES: 263  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/290,665A  
; FILING DATE: 15-AUG-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: RICHARD W. BORK  
; REGISTRATION NUMBER: 36,459  
; REFERENCE/DOCKET NUMBER: 2026-4116  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 758-4800  
; TELEFAX: (212) 751-6849  
; TELEX: 421792  
; INFORMATION FOR SEQ ID NO: 206:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 191 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; ORIGINAL SOURCE:  
; ORGANISM: homoeapiens  
; INDIVIDUAL ISOLATE: HK2  
; US-08-290-665A-206

Query Match 100.0%; Score 1045; DB 2; Length 191;  
Best Local Similarity 100.0%; Pred. No. 9.3e-97;



APPLICANT: MAERTENS, Geert  
TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C  
FILE REFERENCE: 2551-94  
CURRENT APPLICATION NUMBER: US/08/974,690C  
CURRENT FILING DATE: 1997-11-19  
PRIOR APPLICATION NUMBER: PCT/EP94/03555  
PRIOR FILING DATE: 1994-10-28  
PRIOR APPLICATION NUMBER: EP 93402718.6  
PRIOR FILING DATE: 1993-11-04  
NUMBER OF SEQ ID NOS: 286  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 228  
LENGTH: 319  
TYPE: PRT  
ORGANISM: hepatitis C virus  
US-08-974-690C-228

Query Match 100.0%; Score 1045; DB 4; Length 319;  
Best Local Similarity 100.0%; Pred. No. 1.7e-96;  
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSTLPKPKORTKRNTRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRTKTSERSQPRG 60  
DB 1 MSTLPKPKORTKRNTRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRTKTSERSQPRG 60  
QY 61 RRQPIPKARQPGHRWAQPGYWPPLYGNEGCWAGWLLSPRGRPHWGPNDRPRRSRLG 120  
DB 61 RRQPIPKARQPGHRWAQPGYWPPLYGNEGCWAGWLLSPRGRPHWGPNDRPRRSRLG 120  
QY 121 KVIDTLTCGFADLMGYIPVWGAPLGCGVAAALAHGVRAIEDGINATGNLPGCSFSIFLLA 180  
DB 121 KVIDTLTCGFADLMGYIPVWGAPLGCGVAAALAHGVRAIEDGINATGNLPGCSFSIFLLA 180  
QY 181 LLSCLTTPASA 191  
DB 181 LLSCLTTPASA 191

RESULT 5  
US-08-290-665A-191  
Sequence 191, Application US/08290665A  
Patent No. 582852  
GENERAL INFORMATION:  
APPLICANT: BUKH, J., MILLER, R.H. AND  
PURCELL, R.H.  
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND  
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS  
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE  
NUMBER OF SEQUENCES: 263  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/290,665A  
FILING DATE: 15-AUG-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: RICHARD W. BORK  
REGISTRATION NUMBER: 36,459  
REFERENCE/DOCKET NUMBER: 2026-4116  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 191:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 191 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
ORIGINAL SOURCE:  
ORGANISM: homosapiens  
INDIVIDUAL ISOLATE: Z4  
US-08-290-665A-191  
Query Match 95.3%; Score 996; DB 2; Length 191;  
Best Local Similarity 94.8%; Pred. No. 7.3e-92;  
Matches 181; Conservative 4; Mismatches 6; Indels 0; Gaps 0;  
QY 1 MSTLPKPKORTKRNTRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRTKTSERSQPRG 60  
DB 1 MSTLPKPKORTKRNTRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRTKTSERSQPRG 60  
QY 61 RRQPIPKARQPGHRWAQPGYWPPLYGNEGCWAGWLLSPRGRPHWGPNDRPRRSRLG 120  
DB 61 RRQPIPKARQPGHRWAQPGYWPPLYGNEGCWAGWLLSPRGRPHWGPNDRPRRSRLG 120  
QY 121 KVIDTLTCGFADLMGYIPVWGAPLGCGVAAALAHGVRAIEDGINATGNLPGCSFSIFLLA 180  
DB 121 KVIDTLTCGFADLMGYIPVWGAPLGCGVAAALAHGVRAIEDGINATGNLPGCSFSIFLLA 180  
QY 181 LLSCLTTPASA 191  
DB 181 LLSCLTTPASA 191

RESULT 6  
PCT-US95-10398-191  
Sequence 191, Application PC/TUS9510398  
GENERAL INFORMATION:  
APPLICANT: BUKH, J., MILLER, R.H. AND  
PURCELL, R.H.  
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND  
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS  
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE  
NUMBER OF SEQUENCES: 263  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/10398  
FILING DATE: 15-AUG-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/086,428  
FILING DATE: 29 JUNE 1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/290/665  
FILING DATE: 15 AUGUST 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: RICHARD W. BORK  
REGISTRATION NUMBER: 36,459  
REFERENCE/DOCKET NUMBER: 2026-4116

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 191:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 191 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
ORIGINAL SOURCE:  
ORGANISM: homosapiens  
INDIVIDUAL ISOLATE: Z4  
PCT-US95-10398-191

Query Match 95.3%; Score 996; DB 5; Length 191;  
Best Local Similarity 94.8%; Pred. No. 7.3e-92;  
Matches 181; Conservative 4; Mismatches 6; Indels 0; Gaps 0;  
Qy 1 MSTLPKQRTKNTNRPTDVKFPGGQIVGGVYLLPRRGLGVRAIRKTSRSQPRG 60  
Db 1 MSTNPKQRTKNTNRPMQVDFPGGQIVGGVYLLPRRGLGVRAIRKTSRSQPRG 60  
Qy 61 RRQIPKARQPGHWAQPGVPWPLYNCGCGWAGWLLSPRGSRPHGPNDRPRRSNIG 120  
Db 61 RRQIPKARQPGHWAQPGVPWPLYNCGCGWAGWLLSPRGSRPHGPNDRPRRSNIG 120  
Qy 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRRAIEDGINVATGNLPGCSFSIFLLA 180  
Db 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRRAIEDGINVATGNLPGCSFSIFLLA 180  
Qy 181 LLSCLTTPASA 191  
Db 181 LLSCLTTPASA 191

RESULT 7  
US-08-290-665A-192  
Sequence 192, Application US/08290665A  
Patent No. 5882852  
GENERAL INFORMATION:  
APPLICANT: BUKH, J., MILLER, R.H. AND  
APPLICANT: PURCELL, R.H.  
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND  
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS  
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE  
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES  
NUMBER OF SEQUENCES: 263  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/290,665A  
FILING DATE: 15-AUG-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: RICHARD W. BORK  
REGISTRATION NUMBER: 36,459  
REFERENCE/DOCKET NUMBER: 2026-4116  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792

INFORMATION FOR SEQ ID NO: 192:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 191 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
ORIGINAL SOURCE:  
ORGANISM: homosapiens  
INDIVIDUAL ISOLATE: Z8  
US-08-290-665A-192

Query Match 94.0%; Score 982; DB 2; Length 191;  
Best Local Similarity 93.7%; Pred. No. 1.8e-90;  
Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;  
Qy 1 MSTLPKQRTKNTNRPTDVKFPGGQIVGGVYLLPRRGLGVRAIRKTSRSQPRG 60  
Db 1 MSTNPKQRTKNTNRPMQVDFPGGQIVGGVYLLPRRGLGVRAIRKTSRSQPRG 60  
Qy 61 RRQIPKARQPGHWAQPGVPWPLYNCGCGWAGWLLSPRGSRPHGPNDRPRRSNIG 120  
Db 61 RRQIPKARQPGHWAQPGVPWPLYNCGCGWAGWLLSPRGSRPHGPNDRPRRSNIG 120  
Qy 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRRAIEDGINVATGNLPGCSFSIFLLA 180  
Db 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRRAIEDGINVATGNLPGCSFSIFLLA 180  
Qy 181 LLSCLTTPASA 191  
Db 181 LLSCLTTPASA 191

RESULT 8  
US-08-290-665A-193  
Sequence 193, Application US/08290665A  
Patent No. 5882852  
GENERAL INFORMATION:  
APPLICANT: BUKH, J., MILLER, R.H. AND  
APPLICANT: PURCELL, R.H.  
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND  
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS  
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE  
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES  
NUMBER OF SEQUENCES: 263  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/290,665A  
FILING DATE: 15-AUG-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: RICHARD W. BORK  
REGISTRATION NUMBER: 36,459  
REFERENCE/DOCKET NUMBER: 2026-4116  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 193:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 191 amino acids  
TYPE: amino acid

```

; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: Z
US-08-290-665A-193

```

Query Match 94.0%; Score 982; DB 2; Length 191;  
Best Local Similarity 93.7%; Pred. No. 1.8e-90;  
Matches 179: Conservative 5; Mismatches 7; Indels

Qy	1	MSTL	PKQKTKRNTNRPTDVKPPGGQIQVGVYLLPRRGPRLGVRATKTSERSQPRG	60
Db	1	MSTN	PKQKTKRNTNRPMDDVKKPPGGQIQVGVYLLPRRGPRLGVRATKTSERSQPRG	60
Qy	61	RRQI	PKARQPGQRHWAQGYPWPLYGNEGCGAGWLLSPGSRPHWGPNDPRRRSRNLG	120
Db	61	RRQI	PKARSEGSRWAQGYPWPLYGNEGCGAGWLLSPGSRPSWGPNDPRRRSRNLG	120
Qy	121	KVID	TLTCGFADLMGYIPVWGAPLGGVAAALAHGVRAIEDGINYATGNLPGCSFSIFLLA	180
Db	121	KVID	TLTCGFADLMGYIPLVGAPVGVARALAHGVRAVEDGINYATGNLPGCSFSIFLLA	180
Qy	181	LLSCL	TPPASA	191
Db	181	LLSCL	TPPASA	191

## RESULT 9

```

PCT-US95-10398-192
: Sequence 192, Application PC/TUS9510398
: GENERAL INFORMATION:
: APPLICANT: BUXH, J., MILLER, R.H. AND
: APPLICANT: PURCELL, R.H.
: TITLE OF INVENTION: NUCLEOTIDE AND DE
: TITLE OF INVENTION: AMINO ACID SEQUE
: TITLE OF INVENTION: CORE GENES OF ISC
: TITLE OF INVENTION: AND THE USE OF RI
: TITLE OF INVENTION: SEQUENCES IN DIA
: NUMBER OF SEQUENCES: 263
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MORGAN & FINNEGAN
: STREET: 345 PARK AVENUE
: CITY: NEW YORK
: STATE: NEW YORK
: COUNTRY: USA
: ZIP: 10154
: COMPUTER READABLE FORM:
: MEDIUM TYPE: FLOPPY DISK
: COMPUTER: IBM PC COMPATIBLE
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: WORDPERFECT 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/10398
: FILING DATE: 15-AUG-1995
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/086,428
: FILING DATE: 29 JUNE 1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/290/665
: FILING DATE: 15 AUGUST 1994
: ATTORNEY/AGENT INFORMATION:
: NAME: RICHARD W. BORK
: REGISTRATION NUMBER: 36,459
: REFERENCE/DOCKET NUMBER: 2026-4116
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 758-4800
: TELEFAX: (212) 751-6849
: TELEX: 421792
: INFORMATION FOR SEQ ID NO: 192:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 191 amino acids

```

```

; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: Z8
; PCT-US95-10398-192

```

Query Match 94.0%; Score 982; DB 5; Length 191;  
Best Local Similarity 93.7%; Pred. No. 1.8e-90;  
Matches 179; Conservative 5; Mismatches 7; Indels

Qy	1	MSTLKPQKTKRNTNRNRPDVKPPGGQIIVGGHYLLPRGPRLGVZATKTKTSRQPRG	60
Db	1	MSTNPKPQKTKRNTNRNRPMDVKFPGGQIIVGGHYLLPRGPRLGVZATKTKTSRQPRG	60
Qy	61	RQPTPKARQPOGRHWAQPGYPWPLYGNEGCGAGWLLSPRGSRPHWGNPDRRRSRNLG	120
Db	61	RQPTPKARSRGRWAQPGYPWPLYGNEGCGAGWLLSPRGSRPSWGNPDRRSENLG	120
Qy	121	KVTDLTTCGFADLMGYIPVWGAPLGGVAAALAHGVRAIEDGINATGNLPGCSFSIFLLA	180
Db	121	KVTDLTTCGFADLMGYIPLVGAPGVGVARALAHGVRAVEDGINATGNLPGCSFSIFLLA	180
Qy	181	LASCLTTPASA	191
Db	181	LASCLTTPASA	191

## RESULT 10

PC/TUS95-10398-193  
; Sequence 193, Application PC/TUS9510398  
; GENERAL INFORMATION:  
; APPLICANT: BUKH, J., MILLER, R.H. AND  
; APPLICANT: PURCELL, R.H.  
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND  
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS  
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE  
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES  
; NUMBER OF SEQUENCES: 263  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & PINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PC/TUS95/10398  
; FILING DATE: 15-AUG-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/086,428  
; FILING DATE: 29 JUNE 1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/290/665  
; FILING DATE: 15 AUGUST 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: RICHARD W. BORK  
; REGISTRATION NUMBER: 36,459  
; REFERENCE/DOCKET NUMBER: 2026-4116  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 758-4800  
; TELEFAX: (212) 751-6849  
; TELEX: 421792  
; INFORMATION FOR SEQ ID NO: 193:  
; SEQUENCE CHARACTERISTICS:

LENGTH: 191 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
ORIGINAL SOURCE:  
ORGANISM: homosapiens  
INDIVIDUAL ISOLATE: Z1  
PCT-US95-10398-193

Query Match 94.0%; Score 982; DB 5; Length 191;  
Best Local Similarity 93.7%; Pred. No. 1.8e-90;  
Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;  
QY 1 MSTLPKPKQKTKRNTNRPTDVKPPGGQIVGGVYLLPRRGRPLGVRAATKTSRSQPRG 60  
DB 1 MSTNPKPKQKTKRNTNRPTDVKPPGGQIVGGVYLLPRRGRPLGVRAATKTSRSQPRG 60  
QY 61 RQPIPKARQOQGRHWAQPGYPWLYNGEGCGWAGWLLSPRGSPPHMGPNDRSRNLG 120  
DB 61 RQPIPKARQOQGRHWAQPGYPWLYNGEGCGWAGWLLSPRGSPPHMGPNDRSRNLG 120  
QY 121 KVIDTLTCGFADLMGYIPVVGAPLGVAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
DB 121 KVIDTLTCGFADLMGYIPVVGAPLGVAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
QY 181 LLSCLTTPASA 191  
DB 181 LLSCLTTPASA 191

RESULT 11  
US-08-290-665A-172  
Sequence 172, Application US/08290665A  
Patent No. 5882852  
GENERAL INFORMATION:  
APPLICANT: BURKH, J., MILLER, R.H. AND  
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND  
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS  
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE  
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES  
NUMBER OF SEQUENCES: 263  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/290,665A  
FILING DATE: 15-AUG-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: RICHARD W. BORK  
REGISTRATION NUMBER: 36,459  
REFERENCE/DOCKET NUMBER: 2026-4116  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 172:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 191 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown

ORIGINAL SOURCE:  
ORGANISM: homosapiens  
INDIVIDUAL ISOLATE: HK3  
US-08-290-665A-172

Query Match 93.9%; Score 981; DB 2; Length 191;  
Best Local Similarity 93.7%; Pred. No. 2.3e-90;  
Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;  
QY 1 MSTLPKPKQKTKRNTNRPTDVKPPGGQIVGGVYLLPRRGRPLGVRAATKTSRSQPRG 60  
DB 1 MSTNPKPKQKTKRNTNRPTDVKPPGGQIVGGVYLLPRRGRPLGVRAATKTSRSQPRG 60  
QY 61 RQPIPKARQOQGRHWAQPGYPWLYNGEGCGWAGWLLSPRGSPPHMGPNDRSRNLG 120  
DB 61 RQPIPKARQOQGRHWAQPGYPWLYNGEGCGWAGWLLSPRGSPPHMGPNDRSRNLG 120  
QY 121 KVIDTLTCGFADLMGYIPVVGAPLGVAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
DB 121 KVIDTLTCGFADLMGYIPVVGAPLGVAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
QY 181 LLSCLTTPASA 191  
DB 181 LLSCLTTPASA 191

RESULT 12  
PCT-US95-10398-172  
Sequence 172, Application PC/TUS9510398  
GENERAL INFORMATION:  
APPLICANT: BURKH, J., MILLER, R.H. AND  
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND  
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS  
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE  
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES  
NUMBER OF SEQUENCES: 263  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/10398  
FILING DATE: 15-AUG-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/086,428  
FILING DATE: 29 JUNE 1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/290/665  
FILING DATE: 15 AUGUST 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: RICHARD W. BORK  
REGISTRATION NUMBER: 36,459  
REFERENCE/DOCKET NUMBER: 2026-4116  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 172:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 191 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown

```
; TOPOLOGY: unknown
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: HK3
PCT-US95-10398-172

Query Match      93.9%; Score 981; DB 5; Length 191;
Best Local Similarity 93.7%; Pred. No. 2.3e-90;
Matches 179; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKTKRNTNRRPTDVKFFGGQIVGGVILLPRGRGLGVRAATKTSERSQPRG 60
   |||||
Db 1 MSTNPKPKRKTKRNTNRRPMDVKFFGGQIVGGVILLPRGRGLGVRAATKTSERSQPRG 60
   |||||

QY 61 RROQIPKARQPOGRHWAQPGYWPPLYGNEGCGWAGWLLSPRGRPHWGNDPRRRSRNLG 120
   |||||
Db 61 RROQIPKARQPOGRHWAQPGYWPPLYGNEGCGWAGWLLSPRGRPHWGNDPRRRSRNLG 120
   |||||

QY 121 KVIDTLTCGFADLMGYIPVVGAPLGVAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180
   |||||
Db 121 KVIDTLTCGFADLMGYIPVVGAPLGVAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180
   |||||

QY 181 LLSCLTTPASA 191
   |||||
Db 181 LLSCLTTPASA 191
   |||||

RESULT 13
US-08-290-665A-197
; Sequence 197, Application US/08290665A
; Patent No. 5882852
; GENERAL INFORMATION:
; APPLICANT: BUKH, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,665A
; FILING DATE: 15-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 197:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: DK13

US-08-290-665A-197
; Sequence 197, Application PC/TUS9510398
; GENERAL INFORMATION:
; APPLICANT: BUKH, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10398
; FILING DATE: 15-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/086,428
; FILING DATE: 29 JUNE 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/290/665
; FILING DATE: 15 AUGUST 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 197:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
```

INDIVIDUAL ISOLATE: DK13  
PCT-US95-10398-197

Query Match 93.8%; Score 980; DB 5; Length 191;  
Best Local Similarity 93.7%; Pred. No. 2.9e-90;  
Matches 179; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Qy 1 MSTLPKQPKTKRNTNRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60  
Db 1 MSTNPKFQKTKRNTNRPPMDVKFPGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy 61 RRQPIKARQPGRHWAQPGYPWPLYGNEGCGWAGWLLSPRSGRPHWGPNDRRRSRNLG 120  
Db 61 RRQPIKARQLEGRSWAQPGYPWPLYGNEGCGWAGWLLSPRSGRPSWGNDRRRSRNLG 120

Qy 121 KVIDTLTCGFADLMGYIPVVGAPVGGVARALAHGVRLLEDGVNATGNLPGCSFSIFLLA 180  
Db 121 KVIDTLTCGFADLMGYIPVVGAPVGGVARALAHGVRLLEDGVNATGNLPGCSFSIFLLA 180

Qy 181 LLSCLTTPASA 191  
Db 181 LLSCLTTPASA 191

RESULT 15  
US-08-635-886C-217  
; Sequence 217, Application US/08635886C  
; Patent No. 6555114  
; GENERAL INFORMATION:  
; APPLICANT: LEROUX-ROELS, Geert  
; APPLICANT: DELEYS, Robert  
; APPLICANT: MAERTENS, Geert  
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C  
; TITLE OF INVENTION: VIRUS  
; FILE REFERENCE: 2752-18  
; CURRENT APPLICATION NUMBER: US/08/635.886C  
; CURRENT FILING DATE: 1996-04-25  
; PRIOR APPLICATION NUMBER: PCT/EP94/03555  
; PRIOR FILING DATE: 1994-10-28  
; PRIOR APPLICATION NUMBER: EP 93402718.6  
; PRIOR FILING DATE: 1993-11-04  
; NUMBER OF SEQ ID NOS: 286  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 217  
; LENGTH: 319  
; TYPE: PRT  
; ORGANISM: hepatitis C virus  
US-08-635-886C-217

Query Match 93.8%; Score 980; DB 4; Length 319;  
Best Local Similarity 93.7%; Pred. No. 5.5e-90;  
Matches 179; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Qy 1 MSTLPKQPKTKRNTNRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60  
Db 1 MSTNPKFQKTKRNTNRPPMDVKFPGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

Qy 61 RRQPIKARQPGRHWAQPGYPWPLYGNEGCGWAGWLLSPRSGRPHWGPNDRRRSRNLG 120  
Db 61 RRQPIKARQLEGRSWAQPGYPWPLYGNEGCGWAGWLLSPRSGRPSWGNDRRRSRNLG 120

Qy 121 KVIDTLTCGFADLMGYIPVVGAPVGGVARALAHGVRLLEDGVNATGNLPGCSFSIFLLA 180  
Db 121 KVIDTLTCGFADLMGYIPVVGAPVGGVARALAHGVRLLEDGVNATGNLPGCSFSIFLLA 180

Qy 181 LLSCLTTPASA 191  
Db 181 LLSCLTTPASA 191

Search completed: October 30, 2004, 01:47:35  
Job time : 40 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2004, 01:40:17 ; Search time 128 Seconds  
(without alignments)  
483.791 Million cell updates/sec

Title: US-09-084-691B-206  
Perfect score: 1045  
Sequence: 1 MSTLPKQKTKYKRNRRPT.....CSFSIFILLALLSLTTPASA 191

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1370721 seqs, 324215800 residues

Total number of hits satisfying chosen parameters: 1370721

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 20: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1045	100.0	319	15	US-10-651-165-228
2	980	93.8	319	15	US-10-651-165-217
3	977	93.5	319	9	US-09-851-138-42
4	977	93.5	319	15	US-10-651-165-229
5	977	93.5	450	15	US-10-651-165-181
6	977	93.5	2894	9	US-09-941-611-23
7	977	93.5	2894	14	US-10-044-995-23
8	976	93.4	249	15	US-10-365-620-54
9	976	93.4	450	15	US-10-651-165-179
10	976	93.4	450	15	US-10-651-165-180
11	976	93.4	473	15	US-10-365-620-56
12	976	93.4	3011	9	US-09-742-659-4
13	976	93.4	3011	9	US-09-952-572-9

14	976	93.4	3011	9	US-09-929-955-1	Sequence 1, Appli
15	976	93.4	3011	9	US-09-747-419-20	Sequence 20, Appli
16	976	93.4	3011	10	US-09-891-894-3	Sequence 3, Appli
17	976	93.4	3011	13	US-10-104-366-1	Sequence 1, Appli
18	976	93.4	3011	14	US-10-259-275-20	Sequence 20, Appli
19	976	93.4	3011	14	US-10-184-150-3	Sequence 3, Appli
20	976	93.4	3011	14	US-10-328-997-3	Sequence 3, Appli
21	976	93.4	3011	15	US-10-189-359-14	Sequence 14, Appli
22	976	93.4	3011	15	US-10-296-734-406	Sequence 406, App
23	976	93.4	3011	15	US-10-719-619-1	Sequence 1, Appli
24	976	93.4	3012	9	US-09-238-076-2	Sequence 2, Appli
25	976	93.4	3012	10	US-09-995-937-2	Sequence 2, Appli
26	976	93.4	3012	10	US-09-917-563-2	Sequence 2, Appli
27	972	93.0	3011	9	US-09-238-076-20	Sequence 20, Appli
28	972	93.0	3011	10	US-09-995-937-20	Sequence 20, Appli
29	972	93.0	3011	10	US-09-917-563-20	Sequence 20, Appli
30	970	92.8	450	15	US-10-651-165-190	Sequence 190, App
31	969	92.7	450	15	US-10-651-165-189	Sequence 189, App
32	969	92.7	3010	15	US-10-467-000-1	Sequence 1, Appli
33	968	92.6	319	15	US-10-651-165-219	Sequence 219, App
34	968	92.6	3011	9	US-09-916-359-2	Sequence 2, Appli
35	968	92.6	3011	16	US-10-445-724-2	Sequence 3, Appli
36	967	92.5	191	10	US-09-194-949-3	Sequence 187, App
37	965	92.3	450	15	US-10-651-165-187	Sequence 193, App
38	965	92.3	450	15	US-10-651-165-193	Sequence 1, Appli
39	964	92.2	190	14	US-10-268-562-1	Sequence 48, Appli
40	964	92.2	319	9	US-09-851-138-48	Sequence 231, App
41	964	92.2	319	15	US-10-651-165-231	Sequence 7, Appli
42	963	92.2	190	15	US-10-450-649-7	Sequence 50, Appli
43	962	92.1	809	9	US-09-973-025-50	Sequence 50, Appli
44	962	92.1	809	10	US-09-899-303-50	Sequence 50, Appli
45	962	92.1	809	10	US-09-995-808-50	Sequence 50, Appli

## ALIGNMENTS

### RESULT 1

US-10-651-165-228  
; Sequence 228, Application US/10651165  
; Publication No. US2004004787A1  
; GENERAL INFORMATION:  
; APPLICANT: LEROUX-ROELS, Geert  
; APPLICANT: DELEYS, Robert  
; APPLICANT: MAERTENS, Geert  
; TITLE OF INVENTION: VIRUS  
; FILE REFERENCE: 2551-94  
; CURRENT APPLICATION NUMBER: US/10/651,165  
; CURRENT FILING DATE: 2003-09-02  
; PRIOR APPLICATION NUMBER: US/08/974,690C  
; PRIOR FILING DATE: 1997-11-19  
; PRIOR APPLICATION NUMBER: PCT/EP94/03555  
; PRIOR FILING DATE: 1994-10-28  
; PRIOR APPLICATION NUMBER: EP 93402718.6  
; PRIOR FILING DATE: 1993-11-04  
; NUMBER OF SEQ ID NOS: 286  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 228  
; TYPE: PRT  
; ORGANISM: hepatitis C virus  
US-10-651-165-228

Query Match 100.0%; Score 1045; DB 15; Length 319;  
Best Local Similarity 100.0%; Pred. No. 1.le-83;  
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MSTLPKQKTKYKRNRRPTDVKFGGGQIVGGVYLLPRGPGVLGVRAETKTSERQPRG	60
Db	1	MSTLPKQKTKYKRNRRPTDVKFGGGQIVGGVYLLPRGPGVLGVRAETKTSERQPRG	60
Qy	61	RRQPIKARQPCQGRHWAQPGYPWPLYGNECGWAGLLSPRGRPHWPNDRRRNRLG	120

Db 61 RROPIKARQPGRRHWAQPGYPWPLYGNEGCGWAGWLLSPGRSRPHWGPNDRRSRLG 120  
Qy 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
Db 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
Qy 181 LLSCLTTPASA 191  
Db 181 LLSCLTTPASA 191

## RESULT 2

US-10-651-165-217  
; Sequence 217, Application US/10651165  
; Publication No. US20040047877A1  
; GENERAL INFORMATION:  
; APPLICANT: LEROUX-ROELS, Geert  
; APPLICANT: DELEYS, Robert  
; APPLICANT: MAERTENS, Geert  
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C  
; FILE REFERENCE: 2551-94  
; CURRENT APPLICATION NUMBER: US/10/651,165  
; CURRENT FILING DATE: 2003-09-02  
; PRIOR APPLICATION NUMBER: US/08/974,690C  
; PRIOR FILING DATE: 1997-11-19  
; PRIOR APPLICATION NUMBER: PCT/EP94/03555  
; PRIOR FILING DATE: 1994-10-28  
; PRIOR APPLICATION NUMBER: EP 93402718.6  
; PRIOR FILING DATE: 1993-11-04  
; NUMBER OF SEQ ID NOS: 286  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 217  
; LENGTH: 319  
; TYPE: PRT  
; ORGANISM: hepatitis C virus  
US-10-651-165-217

Query Match 93.8%; Score 980; DB 15; Length 319;  
Best Local Similarity 93.7%; Pred. No. 5.9e-78;  
Matches 179; Conservative 4; Mismatches 8; Indels 0; Gaps 0;  
Qy 1 MSTLPKPKQRTKRNTRRPDTVKFPGGQIVGGVYLLPRGPRGLGVRATKTSRSQPRG 60  
Db 1 MSTNPKPKQRTKRNTRRPMDVKFPGGQIVGGVYLLPRGPRGLGVRATKTSRSQPRG 60  
Qy 61 RROPIKARQPGRRHWAQPGYPWPLYGNEGCGWAGWLLSPGRSRPHWGPNDRRSRLG 120  
Db 61 RROPIKARQLEGRSNAQPGYPWPLYGNEGCGWAGWLLSPGRSRPSPWPNDRRSRLG 120  
Qy 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
Db 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
Qy 181 LLSCLTTPASA 191  
Db 181 LLSCLTTPASA 191

## RESULT 3

US-09-851-138-42  
; Sequence 42, Application US/09851138  
; Publication No. US20020183508A1  
; GENERAL INFORMATION:  
; APPLICANT: MAERTENS, GEERT  
; APPLICANT: STUYVER, LIEVEN  
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES  
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC  
; AGENTS  
; NUMBER OF SEQUENCES: 207  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ARNOLD, WHITE & DUNKEE

; STREET: P.O. BOX 4433  
; CITY: HOUSTON  
; STATE: TEXAS  
; COUNTRY: USA  
; ZIP: 77210-4433  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Microsoft Word 6.0 / ASCII text output  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/851,138  
; FILING DATE: 09-May-2001  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/836,075  
; FILING DATE: <Unknown>  
; APPLICATION NUMBER: EP 94870166.9  
; FILING DATE: 21 Oct 1994  
; APPLICATION NUMBER: EP 95870076.7  
; FILING DATE: 28 Jun 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KAMMERER, PATRICIA A.  
; REGISTRATION NUMBER: 29,775  
; REFERENCE/DOCKET NUMBER: INNS:004  
; INFORMATION FOR SEQ ID NO: 42:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 319 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; SEQUENCE DESCRIPTION: SEQ ID NO: 42:  
US-09-851-138-42

Query Match 93.5%; Score 977; DB 9; Length 319;  
Best Local Similarity 93.7%; Pred. No. 1.1e-77;  
Matches 179; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

Qy 1 MSTLPKPKQRTKRNTRRPDTVKFPGGQIVGGVYLLPRGPRGLGVRATKTSRSQPRG 60  
Db 1 MSTLPKPKQRTKRNTRRPMDVKFPGGQIVGGVYLLPRGPRGLGVRATKTSRSQPRG 60  
Qy 61 RROPIKARQPGRRHWAQPGYPWPLYGNEGCGWAGWLLSPGRSRPHWGPNDRRSRLG 120  
Db 61 RROPIKARQPTKRSWQPGYPWPLYGNEGCGWAGWLLSPGRSRPWNPNDRRSRLG 120  
Qy 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
Db 121 KVIDTLTXLADLMGYIPVLGPGVAAALAHGVRAIEDGVNATGNLPGCSFSIFLLA 180  
Qy 181 LLSCLTTPASA 191  
Db 181 LLSCLTTPASA 191

## RESULT 4

US-10-651-165-229  
; Sequence 229, Application US/10651165  
; Publication No. US20040047877A1  
; GENERAL INFORMATION:  
; APPLICANT: LEROUX-ROELS, Geert  
; APPLICANT: DELEYS, Robert  
; APPLICANT: MAERTENS, Geert  
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C  
; FILE REFERENCE: 2551-94  
; CURRENT APPLICATION NUMBER: US/10/651,165  
; CURRENT FILING DATE: 2003-09-02  
; PRIOR APPLICATION NUMBER: US/08/974,690C  
; PRIOR FILING DATE: 1997-11-19  
; PRIOR APPLICATION NUMBER: PCT/EP94/03555  
; PRIOR FILING DATE: 1994-10-28  
; PRIOR APPLICATION NUMBER: EP 93402718.6  
; PRIOR FILING DATE: 1993-11-04

```

; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 229
; LENGTH: 319
; TYPE: PRT
; ORGANISM: hepatitis C virus
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (128)..(129)
; OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (308)..(308)
; OTHER INFORMATION: Xaa is any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (310)..(310)
; OTHER INFORMATION: Xaa is any amino acid
;
US-10-651-165-229

Query Match          93.5%;   Score 977;   DB 15;   Length 319;
Best Local Similarity 93.7%;   Pred. No. 1.le-77;
Matches 179;   Conservative 3;   Mismatches 9;   Indels 0;   Gaps 0;

Qy      1  MSTLPKQRTKRNTRNRPDVKPFGGQIVGGVYLLPRRGRPLGVRATKTSRSQPRG 60
Db      1  MSTLPKQRTKRNTRNRPMDVKPFGGQIVGGVYLLPERRGRPLGVRATKTSRSQPRG 60

Qy      61  RRQPIKARQPGRHWAQPGYPWPPLYNGEGCGWALLSPRGSRPHWGPNDPRRSRLG 120
Db      61  RRQPIKVRQDTGRSWQPGYPWPPLYNGEGCGWALLSPRGSRPNWGPNDPRRSRLG 120

Qy      121  KVIDLTCGFADLMGYIPVVGAPLGSGVAAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180
Db      121  KVIDLTLXLLADLMGYIPVLGSGPLGGVAAALAHGVRAIEDGVNYATGNLPGCSFSIFLLA 180

Qy      181  LLSCLTTPASA 191
Db      181  LLSCLTTPASA 191

```

```

RESULT 5
US-10-651-165-181
; Sequence 181, Application US/10651165
; Publication No. US20040047877A1
; GENERAL INFORMATION:
; APPLICANT: LEROUX-ROELS, Geert
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 181
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus
US-10-651-165-181

```

Query Match 93.5%; Score 977; DB 15; Length 450;  
Best Local Similarity 92.1%; Pred. No. 1.6e-77;  
Matches 176; Conservative 6; Mismatches 9; Indels 0; Gaps 0;  
Ov 1 MSTLPKPKQKRNTRRPTDVKFGGGQIVGGVYLLPRRGRPLGVNTRKTSRSQFRG 60

Db	1	MSITPKPQKTKMTNRRPQDVKPGGGQIVGGVYLLPRGPRLGVRAKTKTSERSQPRG	60
Qy	61	RRQIPKARQPGQHWAQPGVPWPPLYNGEGCGWAGWLLSPGSRPHWGPNDRPRSRNIG	120
Db	61	RRQIPKVRPRPGRTWAQPGVPWPPLYNGEGCGWAGWLLSPGSRPSWGTDPDRSRRNIG	120
Qy	121	KVIDTLTCGGFADLMGYIPIPVVGAPLGGVAAALAHGVRAIETDGINVATGNLPGCSFSIFLLA	180
Db	121	KVIDTLTCGGFADLMGYIPIVVGAPLGGGAARALAHGVRIEDGVNATGNLPGCSFSIFLLA	180
Qy	181	LLSCLTTPASA	191
Db	181	LLSCLTTPASA	191

RESULT 6  
US-09-941-611-23  
; Sequence 23, Application US/09941611  
; Patent No. US20020106640A1  
; GENERAL INFORMATION:  
; APPLICANT: DELEYS, ROBERT J  
; POLLET, DIRK  
; MAERTENS, GEERT  
; VAN HEUVERSWIN, HUGO  
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF  
; ANTIBODIES TO HEPATITIS C VIRUS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHYE P.C.  
; STREET: 1100 NORTH GLEBE ROAD  
; CITY: ARLINGTON  
; STATE: VA  
; COUNTRY: USA

```

ZIF: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30

```

APPLICATION NUMBER: US/09/941,611  
FILING DATE: 30-AUG-2001  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/391,671  
FILING DATE: 1995-02-21  
APPLICATION NUMBER: WO PCT/EP91/02409  
FILING DATE: 13-DEC-1991  
APPLICATION NUMBER: EP 90124241.2  
FILING DATE: 14-DEC-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: SADOFF, B.J.  
REGISTRATION NUMBER: 36,663  
REFERENCE/DOCKET NUMBER: 1487-5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 7038164000  
TELEFAX: 7038164100

INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2894 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
SEQUENCE DESCRIPTION: SEQ ID

S-09-941-611-23

Query Match	93.5%	Score 977;	DB 9;	Length 2894;
Best Local Similarity	92.1%;	Pred. No. 1.2e-76;		
Matches 176;	Conservative	6;	Mismatches	9; Indels



```
; APPLICANT: DELEYS, Robert
; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/10/651,165
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/08/974,690C
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 179
; LENGTH: 450
; TYPE: PRT
; ORGANISM: hepatitis C virus
; US-10-651-165-179

Query Match 93.4%; Score 976; DB 15; Length 450;
Best Local Similarity 92.7%; Pred. No. 1.9e-77;
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKQKTKRNTNRRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATKTSERSQPRG 60
Db 1 MSTNPKPKQKTKRNTNRRPQDVKFPGGQIVGGVYLLPRRGPRLGVRATKTSERSQPRG 60
QY 61 RRQPIPKARQPGRHWAQPGYPWPLYGNEGCGWAGWLLSPGRSRPHWGPNDRRRSRNLG 120
Db 61 RRQPIPKARRPEGRRTWAQPGYPWPLYGNEGCGWAGWLLSPGRSRPSWGPTDPRRSRNLG 120
QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180
Db 121 KVIDTLTCGFADLMGYIPLVGAPLGGVAAALAHGVRAIEDGVVYATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTPASA 191
Db 181 LLSCLTTPASA 191

RESULT 11
US-10-365-620-56
; Sequence 56, Application US/10365620
; Publication No. US20040001853A1
; GENERAL INFORMATION:
; APPLICANT: George, Rajan
; APPLICANT: Tyrrell, Lorne
; APPLICANT: No. US20040001853A1jaim, Antoine
; TITLE OF INVENTION: Chimeric Antigens for Eliciting An Immune Response
; FILE REFERENCE: 656.0016
; CURRENT APPLICATION NUMBER: US/10/365,620
; CURRENT FILING DATE: 2003-02-13
; PRIOR APPLICATION NUMBER: US60/423,578
; PRIOR FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: 60/390,564
; PRIOR FILING DATE: 2002-06-20
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 56
; LENGTH: 473
; TYPE: PRT
; ORGANISM: ORF of HCV Core-TBD protein
; US-10-365-620-56

Query Match 93.4%; Score 976; DB 15; Length 473;
Best Local Similarity 92.7%; Pred. No. 2e-77;
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKQKTKRNTNRRPTDVKFPGGQIVGGVYLLPRRGPRLGVRATKTSERSQPRG 60
Db 31 MSTNPKPKQKTKRNTNRRPQDVKFPGGQIVGGVYLLPRRGPRLGVRATKTSERSQPRG 90
QY 61 RRQPIPKARQPGRHWAQPGYPWPLYGNEGCGWAGWLLSPGRSRPHWGPNDRRRSRNLG 120
Db 91 RRQPIPKARRPEGRRTWAQPGYPWPLYGNEGCGWAGWLLSPGRSRPSWGPTDPRRSRNLG 150
QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180
Db 151 KVIDTLTCGFADLMGYIPLVGAPLGGVAAALAHGVRAIEDGVVYATGNLPGCSFSIFLLA 210
QY 181 LLSCLTTPASA 191
Db 211 LLSCLTTPASA 221

RESULT 12
US-09-742-659-4
; Sequence 4, Application US/09742659
; Patent No. US20010034019A1
; GENERAL INFORMATION:
; APPLICANT: Hong, Zhi
; APPLICANT: Butkiewicz, Nancy J.
; APPLICANT: Zhong, Weidong
; APPLICANT: Ingravallo, Paul
; APPLICANT: Wright-Minogue, Jacquelyn
; APPLICANT: Lau, Johnson Y.
```

; APPLICANT: Lemon, Stanley M.  
; TITLE OF INVENTION: Chimeric HCV/GBV-B viruses  
; FILE REFERENCE: ID01116  
; CURRENT APPLICATION NUMBER: US/09/742,659  
; CURRENT FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: US 60/171,469  
; PRIOR FILING DATE: 1999-12-22  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 3011  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
; US-09-742-659-4

Query Match 93.4%; Score 976; DB 9; Length 3011;  
Best Local Similarity 92.7%; Pred. No. 1.5e-76;  
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;  
QY 1 MSTLPKQKTKRNTNRPTDVKPPGGQIVGGVYLLPRRGPRGLGVRAATKTSERSQPRG 60  
Db 1 MSTNPKPQKTKRNTNRPPQDVKPPGGQIVGGVYLLPRRGPRGLGVRAATKTSERSQPRG 60  
QY 61 RROPIPKARQPGRHWAQPGYPWPLYNCGCGWAGWLLSPRGRPHGPNDRPRRSNRLG 120  
Db 61 RROPIPKARPEGRTWAQPGYPWPLYNCGCGWAGWLLSPRGRSPESWGPTDPRRSNRLG 120  
QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
Db 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
QY 181 LLSCLTTPASA 191  
Db 181 LLSCLTVPASA 191

RESULT 13  
US-09-952-572-9  
; Sequence 9, Application US/09952572  
; Patent No. US20020119495A1  
; GENERAL INFORMATION:  
; APPLICANT: HAWAII BIOTECHNOLOGY GROUP, Inc.  
; APPLICANT: NAKANO, Eileen  
; APPLICANT: CLEMENTS, David  
; APPLICANT: HUMPHREYS, Tom  
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITION OF HEPATITIS C  
; FILE REFERENCE: HAWBIO1100  
; CURRENT APPLICATION NUMBER: US/09/952,572  
; CURRENT FILING DATE: 2001-09-13  
; PRIOR APPLICATION NUMBER: US 60/230,927  
; PRIOR FILING DATE: 2000-09-13  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: Patent In version 3.0  
; SEQ ID NO 9  
; LENGTH: 3011  
; TYPE: PRT  
; ORGANISM: Hepatitis C Virus  
; US-09-952-572-9

Query Match 93.4%; Score 976; DB 9; Length 3011;  
Best Local Similarity 92.7%; Pred. No. 1.5e-76;  
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;  
QY 1 MSTLPKQKTKRNTNRPTDVKPPGGQIVGGVYLLPRRGPRGLGVRAATKTSERSQPRG 60  
Db 1 MSTNPKPQKTKRNTNRPPQDVKPPGGQIVGGVYLLPRRGPRGLGVRAATKTSERSQPRG 60  
QY 61 RROPIPKARQPGRHWAQPGYPWPLYNCGCGWAGWLLSPRGRPHGPNDRPRRSNRLG 120  
Db 61 RROPIPKARPEGRTWAQPGYPWPLYNCGCGWAGWLLSPRGRSPESWGPTDPRRSNRLG 120  
QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
Db 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180

Db 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
QY 181 LLSCLTTPASA 191  
Db 181 LLSCLTVPASA 191

## RESULT 14

US-09-929-955-1  
; Sequence 1, Application US/09929955  
; Patent No. US20020136740A1  
; GENERAL INFORMATION:  
; APPLICANT: Matti Sallberg  
; APPLICANT: Catharina Hultgren  
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND  
; TITLE OF INVENTION: METHODS OF USE THEREOF  
; FILE REFERENCE: TRIPEP.23AUS2  
; CURRENT APPLICATION NUMBER: US/09/929,955  
; CURRENT FILING DATE: 2001-08-15  
; PRIOR APPLICATION NUMBER: 09/705,547  
; PRIOR FILING DATE: 2000-11-03  
; PRIOR APPLICATION NUMBER: 60/229,175  
; PRIOR FILING DATE: 2000-08-29  
; PRIOR APPLICATION NUMBER: 60/225,767  
; PRIOR FILING DATE: 2000-08-17  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 3011  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Hepatitis C virus sequence  
; US-09-929-955-1

Query Match 93.4%; Score 976; DB 9; Length 3011;  
Best Local Similarity 92.7%; Pred. No. 1.5e-76;  
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;  
QY 1 MSTLPKQKTKRNTNRPTDVKPPGGQIVGGVYLLPRRGPRGLGVRAATKTSERSQPRG 60  
Db 1 MSTNPKPQKTKRNTNRPPQDVKPPGGQIVGGVYLLPRRGPRGLGVRAATKTSERSQPRG 60  
QY 61 RROPIPKARQPGRHWAQPGYPWPLYNCGCGWAGWLLSPRGRPHGPNDRPRRSNRLG 120  
Db 61 RROPIPKARPEGRTWAQPGYPWPLYNCGCGWAGWLLSPRGRSPESWGPTDPRRSNRLG 120  
QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
Db 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
QY 181 LLSCLTTPASA 191  
Db 181 LLSCLTVPASA 191

## RESULT 15

US-09-747-419-20  
; Sequence 20, Application US/09747419  
; Patent No. US20020155582A1  
; GENERAL INFORMATION:  
; APPLICANT: Lemon, Stanley  
; APPLICANT: Yi, Minkyung  
; TITLE OF INVENTION: REPLICATION COMPETENT HEPATITIS C VIRUS AND METHODS OF USE  
; FILE REFERENCE: 265.0007 0101  
; CURRENT APPLICATION NUMBER: US/09/747,419  
; CURRENT FILING DATE: 2000-12-23  
; PRIOR APPLICATION NUMBER: US 60/171,909  
; PRIOR FILING DATE: 1999-12-23  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 20  
; LENGTH: 3011

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Polyprotein
US-09-747-419-20

Query Match      93.4%; Score 976; DB 9; Length 3011;
Best Local Similarity 92.7%; Pred.No.1.5e-76;
Matches 177; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSTLPKPKRKTKRNTNRRPTDVKFFGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60
Db 1 MSTNPKPQKTKRNTNRRPDQVKFFGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRG 60

QY 61 RRQPIPKARQPGHWAQGYWPWLYGNEGCGWAGWLLSPGSRPHWGPNDRRSRNLG 120
Db 61 RRQPIPKARRPEGRWAGQGYWPWLYGNEGCGWAGWLLSPGSRPFWGPTDRRSRNLG 120

QY 121 KVIDTLTCGFADLMGYIPVVGAPLGVARAALAHGVRAIEDGINYATGNLPGCSFSIFLLA 180
Db 121 KVIDTLTCGFADLMGYIPLVGAPLGVARAALAHGVRAIEDGVNYATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191
Db 181 LLSCLTVPASA 191
```

Search completed: October 30, 2004, 01:49:51  
Job time : 130 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2004, 01:46:14 ; Search time 69 Seconds

(without alignments)  
993.004 Million cell updates/sec

Title: US-09-084-691B-206

Perfect score: 191

Sequence: 1 MSTLPKQKTKRNTNRPT.....CSFSIFLLALLSCLTTPASA 191

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 2002273 seqs, 358729299 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A.Geneseq 23Sep04:\*

- 1: Genesep1980s:\*
- 2: Genesep1990s:\*
- 3: Genesep2000s:\*
- 4: Genesep2001s:\*
- 5: Genesep2002s:\*
- 6: Genesep2003as:\*
- 7: Genesep2003bs:\*
- 8: Genesep2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	191	100.0	191	2	AAR92987	Hepatitis
2	51	26.7	97	2	AAR66633	Hepatitis
3	51	26.7	97	2	AAW30597	Hepatitis
4	51	26.7	97	2	AAW01624	Protein e
5	51	26.7	97	2	ADf66080	Hepatitis
6	51	26.7	120	5	AAm50828	Hepatitis
7	51	26.7	170	2	AAR29529	HCV core-
8	51	26.7	190	2	AAR92948	Hepatitis
9	51	26.7	191	2	AAR92949	Hepatitis
10	51	26.7	191	2	AAR92983	Hepatitis
11	51	26.7	191	2	AAR92979	Hepatitis
12	51	26.7	191	2	AAR92955	Hepatitis
13	51	26.7	191	2	AAR92980	Hepatitis
14	51	26.7	191	2	AAR92981	Hepatitis
15	51	26.7	191	2	AAR92942	Hepatitis
16	51	26.7	191	2	AAR92986	Hepatitis
17	51	26.7	191	2	AAR92953	Hepatitis
18	51	26.7	191	2	AAR92972	Hepatitis
19	51	26.7	191	2	AAR92982	Hepatitis
20	51	26.7	191	2	AAR92984	Hepatitis
21	51	26.7	246	2	AAW12714	Flu NS1-H
22	51	26.7	411	2	AAR29541	HCV core-
23	51	26.7	412	2	AAR29539	HCV core-
24	51	26.7	470	2	AAR34470	Encoded b
25	51	26.7	470	2	AAR34472	Encoded b

26	51	26.7	470	2	AAR34478	Encoded b
27	51	26.7	470	2	AAR34474	Encoded b
28	51	26.7	470	2	AAR34477	Encoded b
29	51	26.7	470	2	AAR34476	Encoded b
30	51	26.7	470	2	AAR34479	Encoded b
31	51	26.7	470	2	AAR34473	Encoded b
32	51	26.7	470	2	AAR34475	Encoded b
33	51	26.7	470	2	AAR34469	Encoded b
34	51	26.7	470	2	AAR34467	Encoded b
35	51	26.7	470	2	AAR34471	Hepatitis
36	51	26.7	729	2	AAW47263	Cuticle p
37	51	26.7	729	2	AAW47265	Cuticle p
38	51	26.7	1026	8	ADJ82984	Adenovect
39	51	26.7	1031	2	AAR54067	Non-A, no
40	51	26.7	1031	2	AAR98362	5'UTR/COR
41	51	26.7	3010	5	ABG32458	Hepatitis
42	51	26.7	3010	5	ABG32459	Hepatitis
43	51	26.7	3010	5	ABG32451	Hepatitis
44	51	26.7	3010	5	ABG32455	Hepatitis
45	51	26.7	3010	5	ABG32457	Hepatitis

#### ALIGNMENTS

##### RESULT 1

AAR92987  
ID AAR92987 standard; protein; 191 AA.

XX AAR92987;

DT 02-OCT-1996 (first entry)

DE Hepatitis C virus isolate HK2 core protein.

KW HCV; EL; envelope 1; core protein; HCV genotyping; antibody; vaccine;

KW hepatitis.

XX Hepatitis C virus.

OS Hepatitis C virus.

XX WO9605315-A2.

PD 22-FEB-1996.

PF 15-AUG-1995; 95WO-US010398.

PR 15-AUG-1994; 94US-00290665.

XX (USSH ) US SEC DEPT HEALTH.

PI Bukh J, Miller RH, Purcell RH;

DR WPI; 1996-139709/14.

XX N-PSDB; AAT16661.

PT DNA and amino acid sequence of HCV envelope 1 and core proteins - used to determine HCV genotype and as vaccines against HCV infection.

PS Claim 4; Page 223; 340pp; English.

XX AAR92936-R92987 are HCV core proteins derived from 52 different HCV isolates. Isolated cDNA sequences are used for the prodn. of primers useful for detecting the presence of HCV in a sample, the primers are also useful for HCV genotyping. Proteins encoded by the cDNAs can be used in vaccines for immunising against HCV infection. The proteins may also be used to detect antibodies against HCV in serum, saliva, lymphocytes or other mononuclear cells. The antibodies may be used in the prevention of HCV infection

SQ Sequence 191 AA;

Query Match 100.0%; Score 191; DB 2: Length 191;

Best Local Similarity 100.0%; Pred. No. 1.9e-167;



```

AAV01624
ID AAY01624 standard; protein; 97 AA.
XX
AC AAY01624;
XX
DT 21-JUN-1999 (first entry)
XX
DE Protein encoded by HCV J1 untranslated and core domains.
XX
KW HCV; J1 isolate; J1 isolate; HCV1; immunoassay; asiatic strain;
KW diagnosis; HCV infection; blood screening; immunisation; antiviral.
XX
OS Hepatitis C virus.
XX
PN US5871903-A.
XX
PD 16-FEB-1999.
XX
PF 08-MAY-1995; 95US-00436965.
XX
PR 15-SEP-1989; 89US-00408045.
PR 21-DEC-1989; 89US-00456142.
PR 04-JAN-1991; 91US-00637380.
PR 02-AUG-1993; 93US-00101280.
PR 24-FEB-1994; 94US-00201066.
PR 03-NOV-1994; 94US-00334255.
XX
PA (NAHE-) NAT INST OF HEALTH JAPAN.
PA (CHIR ) CHIRON CORP.
XX
PI Saito I, Miyamura T;
DR WPI; 1999-166619/14.
DR N-PSDB; AAX26743.
XX
PT Immunoassays for Asiatic strains of hepatitis C virus - for diagnosis of
PT infection and screening blood supplies.
XX
PS Disclosure; Fig 18; 43pp; English.
XX
CC The present sequence is encoded by the consensus sequence of the
CC untranslated and core domains of a new hepatitis C virus (HCV), J1. The
CC J1 and J7 (also a new HCV isolate) isolates comprise sequences which are
CC distinct from the prototype HCV isolates, HCV1. The specification
CC describes immunoassays for HCV based on antigens from Asiatic strains not
CC cross-reactive with HCV-1. The assays are used for diagnosis of HCV
CC infection and to screen donated blood. The anti-HCV antibodies are also
CC useful therapeutically and prophylactically (passive immunisation); in
CC screening for antiviral agents; for isolation, purification and
CC identification of non-A, non-B hepatitis virus (e.g. by affinity
CC chromatography) and to raise anti-idiotypic antibodies (useful for
CC treatment or diagnosis and to determine immunogenic regions of the HCV
CC antigens)
XX
SQ Sequence 97 AA;
Query Match 26.7%; Score 51; DB 2; Length 97;
Best Local Similarity 100.0%; Pred. No. 8e-39;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKPPGGQIVGGVYLLPRRGLGVRATKTSRSQPRGRQPIPKARQP 71
Db |||||
21 DVKPPGGQIVGGVYLLPRRGLGVRATKTSRSQPRGRQPIPKARQP 71

RESULT 5
ADF66080
ID ADF66080 standard; protein; 97 AA.
XX
AC ADF66080;
XX
DT 12-FEB-2004 (first entry)
XX

Query Match 26.7%; Score 51; DB 2; Length 97;
Best Local Similarity 100.0%; Pred. No. 8e-39;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKPPGGQIVGGVYLLPRRGLGVRATKTSRSQPRGRQPIPKARQP 71
Db |||||
21 DVKPPGGQIVGGVYLLPRRGLGVRATKTSRSQPRGRQPIPKARQP 71

RESULT 6
AAM50828
ID AAM50828 standard; protein; 120 AA.
XX
AC AAM50828;
XX
DT 07-AUG-2003 (revised)
DT 01-MAY-2002 (first entry)
XX
DE Hepatitis C virus core antigen.
XX
KW HCV; core; antigen; immunogen; infection; diagnosis; vaccine.
XX
OS Hepatitis C virus.
XX
FH Key Location/Qualifiers
FT Peptide 1..18
FT Peptide /label= MDL-1
FT Peptide 10..24
FT Peptide /label= MDL-2
FT Peptide 28..42
FT Peptide /label= MDL-3

```

```

DE Hepatitis C virus J1 domain consensus amino acid sequence.
XX
KW Hepatitis C virus; HCV; screening; identification; non A non B hepatitis;
KW NANBH.
XX
OS Hepatitis C virus.
XX
PN US5959092-A.
XX
PD 28-SEP-1999.
XX
PF 08-MAY-1995; 95US-00436966.
XX
PR 15-SEP-1989; 89US-00408045.
PR 21-DEC-1989; 89US-00456142.
PR 04-JAN-1991; 91US-00637380.
PR 02-AUG-1993; 93US-00101280.
PR 24-FEB-1994; 94US-00201066.
PR 03-NOV-1994; 94US-00334255.
XX
PA (CHIR ) CHIRON CORP.
PA (NAHE-) NAT INST OF HEALTH JAPAN.
XX
PI Saito I, Miyamura T;
DR WPI; 1999-570774/48.
DR N-PSDB; ADF66079.
XX
PT Polynucleotide probes useful for screening and identifying carriers of
PT non A non B hepatitis.
XX
PS Example; Fig 18; 44pp; English.
XX
CC The present invention describes polynucleotides sequences from Hepatitis
CC C virus (HCV) isolates J7 C/E, J1 E, J1 E/NS1, J1 NS3, and J1 NS5 or from
CC deposits BP-2593, BP2594, BP-2595, BP-2637 and BP-2638. The
CC polynucleotide sequences are useful as probes for screening and
CC identifying carriers of non A non B hepatitis (NANBH). The present
CC sequence is used in the exemplification of the present invention.
XX
SQ Sequence 97 AA;
Query Match 26.7%; Score 51; DB 2; Length 97;
Best Local Similarity 100.0%; Pred. No. 8e-39;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKPPGGQIVGGVYLLPRRGLGVRATKTSRSQPRGRQPIPKARQP 71
Db |||||
21 DVKPPGGQIVGGVYLLPRRGLGVRATKTSRSQPRGRQPIPKARQP 71

RESULT 6
AAM50828
ID AAM50828 standard; protein; 120 AA.
XX
AC AAM50828;
XX
DT 07-AUG-2003 (revised)
DT 01-MAY-2002 (first entry)
XX
DE Hepatitis C virus core antigen.
XX
KW HCV; core; antigen; immunogen; infection; diagnosis; vaccine.
XX
OS Hepatitis C virus.
XX
FH Key Location/Qualifiers
FT Peptide 1..18
FT Peptide /label= MDL-1
FT Peptide 10..24
FT Peptide /label= MDL-2
FT Peptide 28..42
FT Peptide /label= MDL-3

```

FT Peptide 41..57  
 FT Peptide /label= MDL-4  
 FT Peptide 48..65  
 FT Peptide /label= MDL-5  
 FT Peptide 66..80  
 FT Peptide /label= MDL-6  
 FT Peptide 87..101  
 FT Peptide /label= MDL-7  
 FT Peptide 103..120  
 FT Peptide /label= MDL-8

WO200204484-A2.

17-JAN-2002.

06-JUL-2001; 2001WO-CA000988.

07-JUL-2000; 2000CA-02311022.

(MEDM-) MEDMIRA INC.

Chan HKW, Theolis R;

WPI; 2002-179696/23.

New hepatitis C virus (HCV) peptides and mosaic antigen composition comprising the peptides, useful as immunoreagents for detecting HCV antibodies infection, or as an immunogen for stimulating production of antibodies against HCV.

Disclosure; Page 59; 59pp; English.

The present sequence is that of the core antigen (amino acids 1-120) of the polyprotein encoded by hepatitis C virus (HCV). A panel of overlapping peptides between positions 1 and 186 of the HCV core region was synthesised and tested for immunoreactivity with a commercially available human serum panel. All of the immunoreactivity was concentrated within the region between amino acids 1 to 80. The invention provides a highly immunoreactive mosaic antigen composition (MAC) comprising 2 or more of immunoreactive peptides MDL-1 to MDL-6 (see AAM50816-27) from this region of the core antigen. Each peptide is individually immobilised to a carrier. The unique combination of HCV core peptides in the MAC provides higher specificity and sensitivity for detection of human antibodies specific to HCV in rapid HCV diagnostic applications. An in vitro diagnostic method for detecting anti-HCV antibodies in a test sample and a diagnostic test kit are provided which use the MAC as an immunoreagent. The HCV MAC may also potentially be used as an immunogen in vaccine compositions. (Updated on 07-AUG-2003 to correct OS field.)

SQ Sequence 120 AA;

Query Match 26.7%; Score 51; DB 5; Length 120;  
 Best Local Similarity 100.0%; Pred. No. 9.5e-39;  
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 DVKPPGGQIVGGVYLLPRRGLVTRATKTSRSQPRGRROPPIPKARQP 71  
 Db 21 DVKPPGGQIVGGVYLLPRRGLVTRATKTSRSQPRGRROPPIPKARQP 71

RESULT 7

AAR29529

ID AAR29529 standard; protein; 170 AA.

AC AAR29529;

XX 25-MAR-2003 (revised)

DT 26-APR-1993 (first entry)

XX HCV core-envelope peptide N3-1.

XX Clone; polypeptide; core-envelope; region; Hepatitis C; Virus; HCV; HC;  
 KW cDNA; transcriptase; primer; allele; diagnostic; antibody; suppress;

control; proteolytic processing; precursor.

Hepatitis C virus.

EP518313-A2.

16-DEC-1992.

11-JUN-1992; 92EP-00109812.

11-JUN-1991; 91JP-00139268.

12-JUL-1991; 91JP-00172794.

07-OCT-1991; 91JP-00287008.

16-DEC-1991; 91JP-00332329.

20-APR-1992; 92JP-00099957.

(MITU ) MITSUBISHI KASEI CORP.

Seki M, Honda Y, Takahashi K, Murakami T, Teranishi Y, Hayashi N;

WPI; 1992-417213/51.

N-PSDB; AAQ32438.

New hepatitis C virus gene and its encoded protein - used for diagnosing PT and vaccinating against hepatitis C virus infections.

Claim 5; Page 75-76; 305pp; English.

The sequences given in AAR25929-30 and AAR29534-42 represent various clones of the polypeptide core-envelope region of the Hepatitis C Virus (HCV) gene of the invention. These sequences were isolated from the serum of a patient suffering from hepatitis C (HC). The RNA sequences encoding these peptides were converted into cDNA using transcriptase in the presence of one of the primer sequences given in AAQ32540-46. The cDNA sequences isolated represent different alleles of the same region of the HCV gene. The entire HCV protein (see AAR29527) is useful in the development of a diagnostic method which is more accurate and effective than conventional ones, in the detection of antibodies raised against a wide range of HCVs which have been hardly detected before. The complete gene may be used in an in vitro screening system for a substance capable of specifically suppressing or controlling a proteolytic processing of a precursor polypeptide of HCV. (Updated on 25-MAR-2003 to correct PN field.)

SQ Sequence 170 AA;

Query Match 26.7%; Score 51; DB 2; Length 170;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-38;  
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 DVKPPGGQIVGGVYLLPRRGLVTRATKTSRSQPRGRROPPIPKARQP 71  
 Db 21 DVKPPGGQIVGGVYLLPRRGLVTRATKTSRSQPRGRROPPIPKARQP 71

RESULT 8

AAR29248

ID AAR29248 standard; protein; 190 AA.

XX AAR29248;

XX 02-OCT-1996 (first entry)

XX Hepatitis C virus isolate T10 core protein.

XX HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;  
 KW hepatitis.

XX Hepatitis C virus.

XX WO9605315-A2.

XX 22-FEB-1996.

PD



```
RESULT 11
AAR92979
ID AAR92979 standard; protein; 191 AA.
XX AC AAR92979;
XX DT 02-OCT-1996 (first entry)
XX DE Hepatitis C virus isolate SA4 core protein.
XX KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
XX OS Hepatitis C virus.
XX PN WO9605315-A2.
XX PD 22-FEB-1996.
XX PF 15-AUG-1995; 95WO-US010398.
XX PR 15-AUG-1994; 94US-00290665.
XX PA (USSH ) US SEC DEPT HEALTH.
XX PI Bukh J, Miller RH, Purcell RH;
XX PT DNA and amino acid sequence of HCV envelope 1 and core proteins - used to
XX PT determine HCV genotype and as vaccines against HCV infection.
XX PS Claim 4; Page 216-217; 340pp; English.
XX DR WPI; 1996-139709/14.
XX DR N-PSDB; AAT16653.
XX CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV
XX CC isolates. Isolated cDNA sequences are used for the prodn. of primers
XX CC useful for detecting the presence of HCV in a sample, the primers are
XX CC also useful for HCV genotyping. Proteins encoded by the cDNAs can be used
XX CC in vaccines for immunising against HCV infection. The proteins may also
XX CC be used to detect antibodies against HCV in serum, saliva, lymphocytes or
XX CC other mononuclear cells. The antibodies may be used in the prevention of
XX CC HCV infection
XX SQ Sequence 191 AA;
Query Match 26.7%; Score 51; DB 2; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.4e-38;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 21 DVKPPGGQIVGGVYLLPRGPRGLGVTRATKTSERSQPRGRROPPIKARQP 71
Db 21 DVKPPGGQIVGGVYLLPRGPRGLGVTRATKTSERSQPRGRROPPIKARQP 71
RESULT 12
AAR92955
ID AAR92955 standard; protein; 191 AA.
XX AC AAR92955;
XX DT 02-OCT-1996 (first entry)
XX DE Hepatitis C virus isolate HK4 core protein.
XX KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
XX OS Hepatitis C virus.
XX PN WO9605315-A2.
XX PD 22-FEB-1996.
XX PF 15-AUG-1995; 95WO-US010398.
XX PR 15-AUG-1994; 94US-00290665.
XX PA (USSH ) US SEC DEPT HEALTH.
XX PI Bukh J, Miller RH, Purcell RH;
XX PT DNA and amino acid sequence of HCV envelope 1 and core proteins - used to
XX PT determine HCV genotype and as vaccines against HCV infection.
XX PS Claim 4; Page 217; 340pp; English.
XX DR WPI; 1996-139709/14.
XX DR N-PSDB; AAT16654.
XX CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV
XX CC isolates. Isolated cDNA sequences are used for the prodn. of primers
XX CC useful for detecting the presence of HCV in a sample, the primers are
XX CC also useful for HCV genotyping. Proteins encoded by the cDNAs can be used
XX CC in vaccines for immunising against HCV infection. The proteins may also
XX CC be used to detect antibodies against HCV in serum, saliva, lymphocytes or
XX CC other mononuclear cells. The antibodies may be used in the prevention of
XX CC HCV infection
XX SQ Sequence 191 AA;
Query Match 26.7%; Score 51; DB 2; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.4e-38;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 21 DVKPPGGQIVGGVYLLPRGPRGLGVTRATKTSERSQPRGRROPPIKARQP 71
Db 21 DVKPPGGQIVGGVYLLPRGPRGLGVTRATKTSERSQPRGRROPPIKARQP 71
RESULT 13
AAR92980
ID AAR92980 standard; protein; 191 AA.
XX AC AAR92980;
XX DT 02-OCT-1996 (first entry)
XX DE Hepatitis C virus isolate SA5 core protein.
XX KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
XX OS Hepatitis C virus.
XX PN WO9605315-A2.
XX PD 22-FEB-1996.
XX PF 15-AUG-1995; 95WO-US010398.
XX PR 15-AUG-1994; 94US-00290665.
XX PA (USSH ) US SEC DEPT HEALTH.
XX PI Bukh J, Miller RH, Purcell RH;
XX PT DNA and amino acid sequence of HCV envelope 1 and core proteins - used to
XX PT determine HCV genotype and as vaccines against HCV infection.
XX PS Claim 4; Page 217; 340pp; English.
XX DR WPI; 1996-139709/14.
XX DR N-PSDB; AAT16654.
XX CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV
XX CC isolates. Isolated cDNA sequences are used for the prodn. of primers
XX CC useful for detecting the presence of HCV in a sample, the primers are
XX CC also useful for HCV genotyping. Proteins encoded by the cDNAs can be used
XX CC in vaccines for immunising against HCV infection. The proteins may also
XX CC be used to detect antibodies against HCV in serum, saliva, lymphocytes or
XX CC other mononuclear cells. The antibodies may be used in the prevention of
XX CC HCV infection
XX SQ Sequence 191 AA;
Query Match 26.7%; Score 51; DB 2; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.4e-38;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 21 DVKPPGGQIVGGVYLLPRGPRGLGVTRATKTSERSQPRGRROPPIKARQP 71
Db 21 DVKPPGGQIVGGVYLLPRGPRGLGVTRATKTSERSQPRGRROPPIKARQP 71
```







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2004, 01:49:56 ; Search time 16 Seconds  
(without alignments)  
1148.588 Million cell updates/sec

Title: US-09-084-691B-206

Perfect score: 191

Sequence: 1 MSTLPKQKTKRNTNRPT.....CSFSIFLLALLSCLTPASA 191

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR 79:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	51	26.7	114	2	S41362	genome polyprotein
2	51	26.7	114	2	S41370	genome polyprotein
3	51	26.7	782	2	S18032	genome polyprotein
4	51	26.7	782	2	S18031	genome polyprotein
5	51	26.7	782	2	S19875	genome polyprotein
6	51	26.7	787	2	PN0677	hypothetical prote
7	51	26.7	3010	1	S18030	genome polyprotein
8	50	26.2	112	2	S41371	genome polyprotein
9	50	26.2	114	2	S41365	genome polyprotein
10	49	25.7	82	2	PQ0804	core protein - hep
11	49	25.7	112	2	S41341	genome polyprotein
12	49	25.7	114	2	S41368	genome polyprotein
13	49	25.7	114	2	S41366	genome polyprotein
14	49	25.7	115	2	S41350	genome polyprotein
15	49	25.7	115	2	S41351	genome polyprotein
16	49	25.7	115	2	S41343	genome polyprotein
17	49	25.7	115	2	S41345	genome polyprotein
18	49	25.7	115	2	S41364	genome polyprotein
19	49	25.7	117	2	S41363	genome polyprotein
20	49	25.7	118	2	S41346	genome polyprotein
21	49	25.7	322	2	JN0265	genome polyprotein
22	49	25.7	441	2	S12707	genome polyprotein
23	49	25.7	513	2	A44150	structural protein
24	49	25.7	640	2	JQ1584	genome polyprotein
25	49	25.7	3010	1	GNWVCJ	genome polyprotein
26	49	25.7	3011	1	GNWVC3	genome polyprotein
27	49	25.7	3011	1	GNWVC3	genome polyprotein
28	47	24.6	88	2	S21336	genome polyprotein
29	47	24.6	108	2	S41357	genome polyprotein

## ALIGNMENTS

### RESULT 1

S41362

genome polyprotein - hepatitis C virus (genotype 4, N1) (fragment)

N:Contains: core protein

C:Species: hepatitis C virus

A:Variety: genotype 4, N1

C:Date: 19-May-1994 #sequence\_revision 26-Jul-1996 #text\_change 09-Jul-2004

C:Accession: S41362

R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.

submitted to the EMBL Data Library, January 1994

A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

A:Reference number: S41341

A:Accession: S41362

A:Molecule type: Genomic RNA

A:Residues: 1-114 <VAN>

A:Cross-references: UNIPROT:O68896; EMBL:Z29465; NID:G443892; PIDN:CAA82603.1; PID:G4438

A:Experimental source: genotype 4, N1

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; core protein; polyprotein

F:1-114/Product: core protein #status predicted <MAT>

Query Match

Best Local Similarity 26.7%; Score 51; DB 2; Length 114;

Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKPPGGQIVGGVYLLPRGRLGVRATRTTSERSQPRGRQPIPKARQP 71

DB 21 DVKPPGGQIVGGVYLLPRGRLGVRATRTTSERSQPRGRQPIPKARQP 71

### RESULT 2

S41370

genome polyprotein - hepatitis C virus (genotype 5, N4) (fragment)

N:Contains: core protein

C:Species: hepatitis C virus

A:Variety: genotype 5, N4

C:Date: 19-May-1994 #sequence\_revision 26-Jul-1996 #text\_change 09-Jul-2004

C:Accession: S41370

R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.

submitted to the EMBL Data Library, January 1994

A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

A:Reference number: S41341

A:Accession: S41370

A:Molecule type: Genomic RNA

A:Residues: 1-114 <VAN>

A:Cross-references: UNIPROT:O68904; EMBL:Z29473; NID:G443908; PIDN:CAA82611.1; PID:G4439

A:Experimental source: genotype 5, N4

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; core protein; polyprotein

F:1-114/Product: core protein #status predicted <MAT>

Query Match

Best Local Similarity 26.7%; Score 51; DB 2; Length 114;



RESULT 9

S41365  
genome polyprotein - hepatitis C virus (genotype 4, N4) (fragment)  
N:Contains: core protein  
C:Species: hepatitis C virus  
A:Variety: genotype 4, N4  
C:Date: 19-May-1994 #sequence\_revision 26-Jul-1996 #text\_change 09-Jul-2004  
C:Accession: S41365  
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.  
submitted to the EMBL Data Library, January 1994  
A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.  
A:Reference number: S41341  
A:Accession: S41365  
A:Molecule type: genomic RNA  
A:Residues: 1-114 <VAN>  
A:Cross-references: UNIPROT:Q68899; EMBL:Z29468; NID:G443898; PIDN:CAA82606  
A:Experimental source: genotype 4, N4  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: capsid protein; core protein; polyprotein  
F:1-114/Product: core protein #status predicted <MAT>

Query Match 26.2%; Score 50; DB 2; Length 114;  
Best Local Similarity 100.0%; Pred.No. 7.4e-42;  
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 21 DVKFPGGGQIVGGVLLPRGPRGLGVRATRKTSRSQPRGRQPIPKARQ 70  
|||||  
DB 21 DVKFPGGGQIVGGVLLPRGPRGLGVRATRKTSRSQPRGRQPIPKARQ 70  
|||||

RESULT 10

Q0804  
core protein - hepatitis C virus (fragment)  
C:Species: hepatitis C virus  
C:Date: 03-May-1994 #sequence\_revision 07-Oct-1994 #text\_change 09-Jul-2004  
C:Accession: Q0804; Q0803  
R:Simmonds, P.; McOmish, F.; Yap, P.L.; Chan, S.W.; Lin, C.K.; Dusheiko, G.  
J. Gen. Virol. 74, 661-668, 1993  
A:Title: Sequence variability in the 5' non-coding region of hepatitis C vi  
A:Reference number: Q0803; MUID:93224886; PMID:8385694  
A:Accession: Q0804  
A:Molecule type: mRNA  
A:Residues: 1-82 <STM>  
A:Cross-references: UNIPROT:Q09740; UNIPROT:Q92971; UNIPROT:Q81564; UNIPROT  
PROT:Q8QPS3; UNIPROT:Q8JWL6; UNIPROT:Q8V7S7; UNIPROT:Q8V800; UNIPROT:Q8V700  
B6E2; UNIPROT:Q8B6E1; UNIPROT:Q8B6E0; UNIPROT:Q8B6D7; UNIPROT:Q8B6D6; UNIPROT  
A:Experimental source: variant EG-33 type 4  
A:Accession: Q0803  
A:Molecule type: mRNA  
A:Residues: 1-82 <SI2>  
A:Experimental source: variant EG-29 type 4  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: capsid protein; core protein

Query Match 25.7%; Score 49; DB 2; Length 82;  
Best Local Similarity 100.0%; Pred.No. 5.5e-41;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 21 DVKFPGGGQIVGGVLLPRGPRGLGVRATRKTSRSQPRGRQPIPKAR 69  
|||||  
DB 14 DVKFPGGGQIVGGVLLPRGPRGLGVRATRKTSRSQPRGRQPIPKAR 62  
|||||

RESULT 11

S41341  
genome polyprotein - hepatitis C virus (genotype 1, N1) (fragment)  
N:Contains: core protein  
C:Species: hepatitis C virus  
A:Variety: genotype 1, N1  
C:Date: 19-May-1994 #sequence\_revision 26-Jul-1996 #text\_change 09-Jul-2004  
C:Accession: S41341  
R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.  
submitted to the EMBL Data Library, January 1994

A;Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

A;Reference number: S41341

A;Accession: S41341

A;Molecule type: genomic RNA

A;Residues: 1-112 <VAN>

A;Cross-references: UNIPROT:Q68875; EMBL:Z29444; NID:G443850; PIDN:CAA82582.1; PID:G4438

A;Experimental source: genotype 1, N1

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: capsid protein; core protein; polyprotein

F;1-112/Product: core protein #status predicted <MAT>

Query Match 25.7%; Score 49; DB 2; Length 112;

Best Local Similarity 100.0%; Pred. No. 7.1e-41;

Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKFPGGQIVGGVYLLPRGPRGLGVRATKTSRSQPRGRQPIPKAR 69

DB 21 DVKFPGGQIVGGVYLLPRGPRGLGVRATKTSRSQPRGRQPIPKAR 69

#### RESULT 12

S41368

genome polyprotein - hepatitis C virus (genotype 5, N2) (fragment)

N;Contains: core protein

C;Species: hepatitis C virus

A;Variety: genotype 5, N2

C;Date: 19-May-1994 #sequence\_revision 26-Jul-1996 #text\_change 09-Jul-2004

C;Accession: S41368

R;van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.

submitted to the EMBL Data Library, January 1994

A;Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

A;Reference number: S41341

A;Accession: S41368

A;Molecule type: genomic RNA

A;Residues: 1-114 <VAN>

A;Cross-references: UNIPROT:Q68902; EMBL:Z29471; NID:G443904; PIDN:CAA82609.1; PID:G4439

A;Experimental source: genotype 5, N2

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: capsid protein; core protein; polyprotein

F;1-114/Product: core protein #status predicted <MAT>

Query Match 25.7%; Score 49; DB 2; Length 114;

Best Local Similarity 100.0%; Pred. No. 7.2e-41;

Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKFPGGQIVGGVYLLPRGPRGLGVRATKTSRSQPRGRQPIPKAR 69

DB 21 DVKFPGGQIVGGVYLLPRGPRGLGVRATKTSRSQPRGRQPIPKAR 69

#### RESULT 13

S41366

genome polyprotein - hepatitis C virus (genotype 4, N5) (fragment)

N;Contains: core protein

C;Species: hepatitis C virus

A;Variety: genotype 4, N5

C;Date: 19-May-1994 #sequence\_revision 26-Jul-1996 #text\_change 09-Jul-2004

C;Accession: S41366

R;van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.

submitted to the EMBL Data Library, January 1994

A;Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

A;Reference number: S41341

A;Accession: S41366

A;Molecule type: genomic RNA

A;Residues: 1-114 <VAN>

A;Cross-references: UNIPROT:Q68900; EMBL:Z29469; NID:G443900; PIDN:CAA82607.1; PID:G4439

A;Experimental source: genotype 4, N5

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: capsid protein; core protein; polyprotein

F;1-114/Product: core protein #status predicted <MAT>

Query Match 25.7%; Score 49; DB 2; Length 114;

Best Local Similarity 100.0%; Pred. No. 7.2e-41;

Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKFPGGQIVGGVYLLPRGPRGLGVRATKTSRSQPRGRQPIPKAR 69

DB 21 DVKFPGGQIVGGVYLLPRGPRGLGVRATKTSRSQPRGRQPIPKAR 69

#### RESULT 14

S41350

genome polyprotein - hepatitis C virus (genotype 1, N8) (fragment)

N;Contains: core protein

C;Species: hepatitis C virus

A;Variety: genotype 1, N8

C;Date: 19-May-1994 #sequence\_revision 26-Jul-1996 #text\_change 09-Jul-2004

C;Accession: S41350

R;van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.

submitted to the EMBL Data Library, January 1994

A;Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

A;Reference number: S41341

A;Accession: S41350

A;Molecule type: genomic RNA

A;Residues: 1-115 <VAN>

A;Cross-references: UNIPROT:Q68884; EMBL:Z29453; NID:G443868; PIDN:CAA82591.1; PID:G4438

A;Experimental source: genotype 1, N8

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: capsid protein; core protein; polyprotein

F;1-115/Product: core protein #status predicted <MAT>

Query Match 25.7%; Score 49; DB 2; Length 115;

Best Local Similarity 100.0%; Pred. No. 7.3e-41;

Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKFPGGQIVGGVYLLPRGPRGLGVRATKTSRSQPRGRQPIPKAR 69

DB 21 DVKFPGGQIVGGVYLLPRGPRGLGVRATKTSRSQPRGRQPIPKAR 69

#### RESULT 15

S41351

genome polyprotein - hepatitis C virus (genotype 1, N9) (fragment)

N;Contains: core protein

C;Species: hepatitis C virus

A;Variety: genotype 1, N9

C;Date: 19-May-1994 #sequence\_revision 26-Jul-1996 #text\_change 09-Jul-2004

C;Accession: S41351

R;van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.

submitted to the EMBL Data Library, January 1994

A;Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

A;Reference number: S41341

A;Accession: S41351

A;Molecule type: genomic RNA

A;Residues: 1-115 <VAN>

A;Cross-references: UNIPROT:Q68885; EMBL:Z29454

A;Experimental source: genotype 1, N9

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: capsid protein; core protein; polyprotein

F;1-115/Product: core protein #status predicted <MAT>

Query Match 25.7%; Score 49; DB 2; Length 115;

Best Local Similarity 100.0%; Pred. No. 7.3e-41;

Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKFPGGQIVGGVYLLPRGPRGLGVRATKTSRSQPRGRQPIPKAR 69

DB 21 DVKFPGGQIVGGVYLLPRGPRGLGVRATKTSRSQPRGRQPIPKAR 69

Search completed: October 30, 2004, 01:58:16

Job time : 16 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2004, 01:52:44 ; Search time 67 Seconds  
(without alignments)  
1640.247 Million cell updates/sec

Title: US-09-084-691B-206

Perfect score: 191  
Sequence: 1 MSTLPKPKQKTKRNTNRPT.....CSFSIFLLALLSCLTTPASA 191

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0  
Searched: 1825181 seqs, 575374646 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : UniProt 02: \*  
1: uniprot\_sprot: \*  
2: uniprot\_trembl: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	191	100.0	191	2	Q68115	Q68115 hepatitis c
2	161	84.3	191	2	Q68411	Q68411 hepatitis c
3	116	60.7	414	2	Q81270	Q81270 hepatitis c
4	106	55.5	414	2	Q81269	Q81269 hepatitis c
5	97	50.8	117	2	Q75Q30	Q75Q30 hepatitis c
6	97	50.8	117	2	Q75Q31	Q75Q31 hepatitis c
7	97	50.8	117	2	Q75Q33	Q75Q33 hepatitis c
8	97	50.8	117	2	BAD11958	BAD11958 hepatitis c
9	97	50.8	117	2	BAD11960	BAD11960 hepatitis c
10	97	50.8	117	2	BAD11961	BAD11961 hepatitis c
11	86	45.0	414	2	P89956	P89956 hepatitis c
12	85	44.5	414	2	Q81268	Q81268 hepatitis c
13	85	44.5	414	2	Q81329	Q81329 hepatitis c
14	85	44.5	3016	2	O92531	O92531 hepatitis c
15	82	42.9	414	2	Q81271	Q81271 hepatitis c
16	81	42.4	414	2	Q81314	Q81314 hepatitis c
17	79	41.4	319	2	Q68703	Q68703 hepatitis c
18	79	41.4	414	2	P89962	P89962 hepatitis c
19	79	41.4	414	2	P89963	P89963 hepatitis c
20	73	38.2	109	2	Q81493	Q81493 hepatitis c
21	67	35.1	158	2	Q80830	Q80830 hepatitis c
22	65	34.0	319	2	Q68690	Q68690 hepatitis c
23	63	33.0	109	2	Q81494	Q81494 hepatitis c
24	63	33.0	415	2	P87751	P87751 hepatitis c
25	63	33.0	3013	2	O92530	O92530 hepatitis c
26	56	29.3	3018	2	O39927	O39927 hepatitis c
27	55	28.8	319	2	Q68705	Q68705 hepatitis c
28	53	27.7	117	2	Q75Q32	Q75Q32 hepatitis c
29	53	27.7	117	2	BAD11959	BAD11959 hepatitis c
30	53	27.7	414	2	P89958	P89958 hepatitis c
31	51	26.7	74	2	Q68708	Q68708 hepatitis c

32 51 26.7 83 2 Q81264 hepatitis c  
33 51 26.7 83 2 Q97980 hepatitis c  
34 51 26.7 85 2 Q91KF2 hepatitis c  
35 51 26.7 85 2 Q91KF3 hepatitis c  
36 51 26.7 85 2 Q91KF4 hepatitis c  
37 51 26.7 85 2 Q91KF7 hepatitis c  
38 51 26.7 85 2 Q91KG2 hepatitis c  
39 51 26.7 85 2 Q91KG3 hepatitis c  
40 51 26.7 85 2 Q91KG5 hepatitis c  
41 51 26.7 85 2 Q91KG9 hepatitis c  
42 51 26.7 85 2 Q91KH0 hepatitis c  
43 51 26.7 85 2 Q91KH3 hepatitis c  
44 51 26.7 85 2 Q91KH4 hepatitis c  
45 51 26.7 85 2 Q91KH5 hepatitis c

#### ALIGNMENTS

##### RESULT 1

Q68115 PRELIMINARY: PRT; 191 AA.  
AC Q68115; 01-NOV-1996 (TREMELrel. 01, Created)  
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
DE Polypeptide (fragment).  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94336721; PubMed=8058787;  
RA Bukh J., Purcell R.H., Miller R.H.;  
RT "Sequence analysis of the core gene of 14 hepatitis C virus  
genotypes";  
RL Proc. Natl. Acad. Sci. U.S.A. 91:8239-8243(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92279243; PubMed=1317578;  
RA Bukh J., Purcell R.H., Miller R.H.;  
RT "Sequence analysis of the 5' noncoding region of hepatitis C virus";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:4942-4946(1992).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93376778; PubMed=8396266;  
RA Bukh J., Purcell R.H., Miller R.H.;  
RT "At least 12 genotypes of hepatitis C virus predicted by sequence  
analysis of the putative E1 gene of isolates collected worldwide";  
RL Proc. Natl. Acad. Sci. U.S.A. 90:8234-8238(1993).  
DR EMBL; U0198; AAA21037.1; -  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0005198; P:structural molecule activity; IEA.  
DR InterPro; IPR002522; HCV capsid.  
DR InterPro; IPR002521; HCV core.  
DR Pfam; PF01543; HCV\_capsid; 1.  
DR Polyprotein.  
FT CHAIN 1 >191 core protein.  
FT NON TER 191 191  
SQ SEQUENCE 191 AA; 20725 MW; 7A7994DF2F0909C6 CRC64;

Query Match 100.0%; Score 191; DB 2; Length 191;  
Best Local Similarity 100.0%; Pred. No. 1.8e-179;  
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSTLPKPKQKTKRNTNRPTDKPFGGQIYGVYLLPRGPRIGVTRTKTSERQ 60  
DB 1 MSTLPKPKQKTKRNTNRPTDKPFGGQIYGVYLLPRGPRIGVTRTKTSERQ 60  
QY 61 RROPTPKARQPGGRHWAQPGYPWPLYGNCGGAGLLSPGSRPHWGPNDRRRSRNLG 120

```

Db 61 RRQPIKARQPGQHWAQPGYPWPVLYNECGWAGWLLSPRGRPHWPNDRRRSRNLG 120
QY 121 KVDTLTGCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGYNATGNLPGCSFSIFLLA 180
Db 121 KVDTLTGCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGYNATGNLPGCSFSIFLLA 180
QY 181 LLSCLTTTPASA 191
Db 181 LLSCLTTTPASA 191

RESULT 2
Q68411 PRELIMINARY; PRT; 191 AA.
AC Q68411; 191 AA.
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Core protein (Fragment).
OS Hepatitis C virus type 6.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=42182;
RN [1]
RP SEQUENCE FROM N.A.
RA Barnier L., Willems B., Delage G., Murphy D.G.;
RT "Identification of numerous hepatitis C virus genotypes in Montreal,
RT Canada.";
RL J. Clin. Microbiol. 34:2815-2818 (1996).
DR EMBL; U33435; AAB40038.1; -.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
FT NON_TER 191_191
QY SEQUENCE 191 AA; 20789 MW; F292AB64B56DE30A CRC64;

Query Match 84.3%; Score 161; DB 2; Length 191;
Best Local Similarity 100.0%; Pred. No. 6.4e-150;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKFPGGQVGGVYLLPRGGRLGVATRKTSERSQPRGRORIPKARQPGHWAQPG 80
Db 21 DVKFPGGQVGGVYLLPRGGRLGVATRKTSERSQPRGRORIPKARQPGHWAQPG 80
QY 81 YPWPVLYNECGWAGWLLSPRGRPHWPNDRRRSRNLGKVIDTLTCGFADLMGYIPV 140
Db 81 YPWPVLYNECGWAGWLLSPRGRPHWPNDRRRSRNLGKVIDTLTCGFADLMGYIPV 140
QY 141 GAPIGGVAAALAHGVRAIEDGYNATGNLPGCSFSIFLLAL 181
Db 141 GAPIGGVAAALAHGVRAIEDGYNATGNLPGCSFSIFLLAL 181

RESULT 3
Q81270 PRELIMINARY; PRT; 414 AA.
AC Q81270;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Core, env and part of E2/NS1 (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Tokita H., Okamoto H., Luengrojanakul P., Vareesangthip K.,
RX MEDLINE=96005057; PubMed=7561773;
RA Chainuvati T., Iizuka H., Tsuda F., Miyakawa Y., Mayumi M.;
RT "Hepatitis C virus variants from Thailand classifiable into five novel
RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
RT genetic groups.";
RL J. Gen. Virol. 76:2329-2335 (1995).
DR EMBL; D37844; BAA07090.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 414
QY SEQUENCE 414 AA; 44994 MW; 8C1E123FC61D89FC CRC64;

Query Match 60.7%; Score 116; DB 2; Length 414;
Best Local Similarity 100.0%; Pred. No. 2.5e-105;
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 WAQPGYPWPVLYNECGWAGWLLSPRGRPHWPNDRRRSRNLGKVIDTLTCGFADLMG 135
Db 76 WAQPGYPWPVLYNECGWAGWLLSPRGRPHWPNDRRRSRNLGKVIDTLTCGFADLMG 135
QY 136 YIPVVGAPLGGVAAALAHGVRAIEDGYNATGNLPGCSFSIFLLALSLTTTPASA 191
Db 136 YIPVVGAPLGGVAAALAHGVRAIEDGYNATGNLPGCSFSIFLLALSLTTTPASA 191

RESULT 4
Q81269 PRELIMINARY; PRT; 414 AA.
AC Q81269;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Core, env and part of E2/NS1 (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Tokita H., Okamoto H., Luengrojanakul P., Vareesangthip K.,
RX MEDLINE=96005057; PubMed=7561773;
RA Chainuvati T., Iizuka H., Tsuda F., Miyakawa Y., Mayumi M.;
RT "Hepatitis C virus variants from Thailand classifiable into five novel
RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
RT genetic groups.";
RL J. Gen. Virol. 76:2329-2335 (1995).
DR EMBL; D37844; BAA07090.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 414
QY SEQUENCE 414 AA; 44994 MW; 8C1E123FC61D89FC CRC64;

```

```

RA Chainuvati T., Iizuka H., Tsuda F., Miyakawa Y., Mayumi M.;
RT "Hepatitis C virus variants from Thailand classifiable into five novel
RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
RT genetic groups.";
RL J. Gen. Virol. 76:2329-2335 (1995).
DR EMBL; D37845; BAA07091.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 414
QY SEQUENCE 414 AA; 44994 MW; 8C1E123FC61D89FC CRC64;

Query Match 60.7%; Score 116; DB 2; Length 414;
Best Local Similarity 100.0%; Pred. No. 2.5e-105;
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 WAQPGYPWPVLYNECGWAGWLLSPRGRPHWPNDRRRSRNLGKVIDTLTCGFADLMG 135
Db 76 WAQPGYPWPVLYNECGWAGWLLSPRGRPHWPNDRRRSRNLGKVIDTLTCGFADLMG 135
QY 136 YIPVVGAPLGGVAAALAHGVRAIEDGYNATGNLPGCSFSIFLLALSLTTTPASA 191
Db 136 YIPVVGAPLGGVAAALAHGVRAIEDGYNATGNLPGCSFSIFLLALSLTTTPASA 191

RESULT 4
Q81269 PRELIMINARY; PRT; 414 AA.
AC Q81269;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Core, env and part of E2/NS1 (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Tokita H., Okamoto H., Luengrojanakul P., Vareesangthip K.,
RX MEDLINE=96005057; PubMed=7561773;
RA Chainuvati T., Iizuka H., Tsuda F., Miyakawa Y., Mayumi M.;
RT "Hepatitis C virus variants from Thailand classifiable into five novel
RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
RT genetic groups.";
RL J. Gen. Virol. 76:2329-2335 (1995).
DR EMBL; D37844; BAA07090.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 414
QY SEQUENCE 414 AA; 44994 MW; 8C1E123FC61D89FC CRC64;

```

```
SQ SEQUENCE 414 AA; 45072 MW; 6C9B1AFE4642241D CRC64;

Query Match      55.5%; Score 106; DB 2; Length 414;
Best Local Similarity 100.0%; Pred. No. 1.7e-95;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 WAQPGYPMFLYGNCGCGWAGWLLSPRGSRPHWGPNDRPRRSRLGKVIDTLTTCGFADLMG 135
Db 76 WAQPGYPMFLYGNCGCGWAGWLLSPRGSRPHWGPNDRPRRSRLGKVIDTLTTCGFADLMG 135

QY 136 YIPVVGAPLGGVAAALAHGVRATEDGINATGNLPCGSFIFLLAL 181
Db 136 YIPVVGAPLGGVAAALAHGVRATEDGINATGNLPCGSFIFLLAL 181

RESULT 5
Q75Q30
ID Q75Q30 PRELIMINARY; PRT; 117 AA.
AC Q75Q30;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DE Polyprotein (Fragment).
OS Hepatitis C virus type 6a.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus; Hepatitis C virus type 6.
OX NCBI_TaxID=31655;
RN [1]_
RP SEQUENCE FROM N.A.
RA Hirano M., Tran H.T., Abe K.;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB162867; BAD11961.1; -.
DR InterPro; IPR002522; HCV capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER 117
SQ SEQUENCE 117 AA; 13441 MW; 9EA99F0E118A3F20 CRC64;

Query Match      50.8%; Score 97; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 4.6e-87;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKFFGGGQIVGGVYLLPRRGPRLGVRAATKTSERSQPRGRQPIPKARQPOGRHWAQPG 80
Db 21 DVKFFGGGQIVGGVYLLPRRGPRLGVRAATKTSERSQPRGRQPIPKARQPOGRHWAQPG 80

QY 81 YPWPLYGNEGCCWAGWLLSPRGSRPHWGPNDRPRRSR 117
Db 81 YPWPLYGNEGCCWAGWLLSPRGSRPHWGPNDRPRRSR 117

RESULT 6
Q75Q31
ID Q75Q31 PRELIMINARY; PRT; 117 AA.
AC Q75Q31;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DE Polyprotein (Fragment).
OS Hepatitis C virus type 6a.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus; Hepatitis C virus type 6.
OX NCBI_TaxID=31655;
RN [1]_
RP SEQUENCE FROM N.A.
RA Hirano M., Tran H.T., Abe K.;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB162866; BAD11960.1; -.
DR InterPro; IPR002522; HCV capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER 117
SQ SEQUENCE 117 AA; 13407 MW; 3EAE29D8CAE789F6 CRC64;

Query Match      50.8%; Score 97; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 4.6e-87;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKFFGGGQIVGGVYLLPRRGPRLGVRAATKTSERSQPRGRQPIPKARQPOGRHWAQPG 80
Db 21 DVKFFGGGQIVGGVYLLPRRGPRLGVRAATKTSERSQPRGRQPIPKARQPOGRHWAQPG 80

QY 81 YPWPLYGNEGCCWAGWLLSPRGSRPHWGPNDRPRRSR 117
Db 81 YPWPLYGNEGCCWAGWLLSPRGSRPHWGPNDRPRRSR 117

RESULT 7
Q75Q33
ID Q75Q33 PRELIMINARY; PRT; 117 AA.
AC Q75Q33;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DE Polyprotein (Fragment).
OS Hepatitis C virus type 6a.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus; Hepatitis C virus type 6.
OX NCBI_TaxID=31655;
RN [1]_
RP SEQUENCE FROM N.A.
RA Hirano M., Tran H.T., Abe K.;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB162864; BAD11958.1; -.
DR InterPro; IPR002522; HCV capsid.
DR Pfam; PF01543; HCV_capsid; 1.
KW Polyprotein.
FT NON_TER 117
SQ SEQUENCE 117 AA; 13407 MW; 3EAE29D8CAE789F6 CRC64;

Query Match      50.8%; Score 97; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 4.6e-87;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKFFGGGQIVGGVYLLPRRGPRLGVRAATKTSERSQPRGRQPIPKARQPOGRHWAQPG 80
Db 21 DVKFFGGGQIVGGVYLLPRRGPRLGVRAATKTSERSQPRGRQPIPKARQPOGRHWAQPG 80

QY 81 YPWPLYGNEGCCWAGWLLSPRGSRPHWGPNDRPRRSR 117
Db 81 YPWPLYGNEGCCWAGWLLSPRGSRPHWGPNDRPRRSR 117

RESULT 8
BAD11958
ID BAD11958 PRELIMINARY; PRT; 117 AA.
AC BAD11958;
DT 03-MAR-2004 (TReMBLrel. 27, Created)
DT 03-MAR-2004 (TReMBLrel. 27, Last sequence update)
DE Polyprotein (Fragment).
OS Hepatitis C virus type 6a.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus; Hepatitis C virus type 6.
OX NCBI_TaxID=31655;
RN [1]_
RP SEQUENCE FROM N.A.
RA Hirano M., Tran H.T., Abe K.;
RT "Genotypic distribution of hepatitis C virus (HCV) in Ho Chi Minh City, Vietnam: New genotyping systems for identification of Vietnamese HCV isolates.";
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB162864; BAD11958.1; -.
KW Polyprotein.
FT NON_TER 117
```

```
SQ SEQUENCE 117 AA; 13407 MW; 3EAE29D8CAE789F6 CRC64;
Query Match 50.8%; Score 97; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 4.6e-87;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKFPGGQIVGGVYLLPRGPRILGVTRATKTSERSQPRGRROPPIPKAROPQGRHWAQPG 80
DB 21 DVKFPGGQIVGGVYLLPRGPRILGVTRATKTSERSQPRGRROPPIPKAROPQGRHWAQPG 80
QY 81 YPWPLYNCGCGWAGWLLSPRGSRPHWGPNDPRRSR 117
DB 81 YPWPLYNCGCGWAGWLLSPRGSRPHWGPNDPRRSR 117

RESULT 9
BAD11960
ID BAD11960 PRELIMINARY; PRT; 117 AA.
AC BAD11960;
DT 03-MAR-2004 (TREMBlrel. 27, Created)
DT 03-MAR-2004 (TREMBlrel. 27, Last sequence update)
DE Polypotein (Fragment).
OS Hepatitis C virus type 6a.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus; Hepatitis C virus; Hepatitis C virus type 6.
OX NCBI_TaxID=31655;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VT657;
RA Hirano M., Tran H.T., Abe K.;
RT "Genotypic distribution of hepatitis C virus (HCV) in Ho Chi Minh City, Vietnam: New genotyping systems for identification of Vietnamese HCV isolates."
RT Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB162866; BAD11960.1; -.
KW Polypotein.
FT NON_TER 117 117
SQ SEQUENCE 117 AA; 13407 MW; 3EAE29D8CAE789F6 CRC64;

Query Match 50.8%; Score 97; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 4.6e-87;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKFPGGQIVGGVYLLPRGPRILGVTRATKTSERSQPRGRROPPIPKAROPQGRHWAQPG 80
DB 21 DVKFPGGQIVGGVYLLPRGPRILGVTRATKTSERSQPRGRROPPIPKAROPQGRHWAQPG 80
QY 81 YPWPLYNCGCGWAGWLLSPRGSRPHWGPNDPRRSR 117
DB 81 YPWPLYNCGCGWAGWLLSPRGSRPHWGPNDPRRSR 117

RESULT 11
P89956
ID P89956 PRELIMINARY; PRT; 414 AA.
AC P89956;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Core, env and part of E2/NS1 (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
MEDLINE=95062197; PubMed=7972001;
RA Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T., Izuka H., Mishiro S., Miyakawa Y., Mayumi M.;
RT "Hepatitis C virus variants from Vietnam are classifiable into the seventh, eighth, and ninth major genetic groups."
RL Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026(1994).
RN [2]
RP SEQUENCE FROM N.A.
OKamoto H.;
RA Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RL EMBL; D88469; BAA13618.1; -.
DR PIR; PQ0804; PQ0804.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
FT NON_TER 414 414
SQ SEQUENCE 414 AA; 44528 MW; 7A3BE1710311C017 CRC64;

Query Match 45.0%; Score 86; DB 2; Length 414;
Best Local Similarity 100.0%; Pred. No. 8.6e-76;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 QGRHWAQGYWPVLYNCGCGWAGWLLSPRGSRPHWGPNDPRRSRNLGKVIDTILTCGFA 131
DB 72 QGRHWAQGYWPVLYNCGCGWAGWLLSPRGSRPHWGPNDPRRSRNLGKVIDTILTCGFA 131
QY 132 DLMGYIPVVGAPLGGVAAALAHGVRA 157
DB 132 DLMGYIPVVGAPLGGVAAALAHGVRA 157
```

```
SQ SEQUENCE 117 AA; 13407 MW; 3EAE29D8CAE789F6 CRC64;
Query Match 50.8%; Score 97; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 4.6e-87;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKFPGGQIVGGVYLLPRGPRILGVTRATKTSERSQPRGRROPPIPKAROPQGRHWAQPG 80
DB 21 DVKFPGGQIVGGVYLLPRGPRILGVTRATKTSERSQPRGRROPPIPKAROPQGRHWAQPG 80
QY 81 YPWPLYNCGCGWAGWLLSPRGSRPHWGPNDPRRSR 117
DB 81 YPWPLYNCGCGWAGWLLSPRGSRPHWGPNDPRRSR 117

RESULT 9
BAD11960
ID BAD11960 PRELIMINARY; PRT; 117 AA.
AC BAD11960;
DT 03-MAR-2004 (TREMBlrel. 27, Created)
DT 03-MAR-2004 (TREMBlrel. 27, Last sequence update)
DE Polypotein (Fragment).
OS Hepatitis C virus type 6a.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus; Hepatitis C virus; Hepatitis C virus type 6.
OX NCBI_TaxID=31655;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VT657;
RA Hirano M., Tran H.T., Abe K.;
RT "Genotypic distribution of hepatitis C virus (HCV) in Ho Chi Minh City, Vietnam: New genotyping systems for identification of Vietnamese HCV isolates."
RT Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB162866; BAD11960.1; -.
KW Polypotein.
FT NON_TER 117 117
SQ SEQUENCE 117 AA; 13407 MW; 3EAE29D8CAE789F6 CRC64;

Query Match 50.8%; Score 97; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 4.6e-87;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKFPGGQIVGGVYLLPRGPRILGVTRATKTSERSQPRGRROPPIPKAROPQGRHWAQPG 80
DB 21 DVKFPGGQIVGGVYLLPRGPRILGVTRATKTSERSQPRGRROPPIPKAROPQGRHWAQPG 80
QY 81 YPWPLYNCGCGWAGWLLSPRGSRPHWGPNDPRRSR 117
DB 81 YPWPLYNCGCGWAGWLLSPRGSRPHWGPNDPRRSR 117

RESULT 10
BAD11961
ID BAD11961 PRELIMINARY; PRT; 117 AA.
AC BAD11961;
DT 03-MAR-2004 (TREMBlrel. 27, Created)
DT 03-MAR-2004 (TREMBlrel. 27, Last sequence update)
DE Polypotein (Fragment).
OS Hepatitis C virus type 6a.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus; Hepatitis C virus; Hepatitis C virus type 6.
OX NCBI_TaxID=31655;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VT689;
RA Hirano M., Tran H.T., Abe K.;
RT "Genotypic distribution of hepatitis C virus (HCV) in Ho Chi Minh City, Vietnam: New genotyping systems for identification of Vietnamese HCV isolates."
RT Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
RL
```



```
RESULT 12
Q81268
ID Q81268 PRELIMINARY; PRT; 414 AA.
AC Q81268;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Core, env, and part of E2/NS1 (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96005057; PubMed=7561773;
RA Tokita H., Okamoto H., Luengrojjanakul P., Vareesangthip K.,
RA Chaiyavati T., Iizuka H., Tsuda F., Miyakawa Y., Mayumi M.;
RT "Hepatitis C virus variants from Thailand classifiable into five novel
RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major
RT genetic groups.";
RL J. Gen. Virol. 76:2329-2335 (1995).
DR EMBL; D37843; BAA07089.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; P:structural molecule activity; IEA.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 414
SQ SEQUENCE 414 AA; 44836 MW; 212740491A9DA0B1 CRC64;

Query Match 44.5%; Score 85; DB 2; Length 414;
Best Local Similarity 100.0%; Pred. No. 8.4e-75;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 WGNPDRRRNLGKVIDITLCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVAT 166
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 167 GNLPCCSFIFLLALLSCLITTPASA 191
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 13
Q81329
ID Q81329 PRELIMINARY; PRT; 414 AA.
AC Q81329;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Core, env and part of E2/NS1 (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95062197; PubMed=7972001;
RA Tokita H., Okamoto H., Tsuda F., Song P., Nakata S., Chosa T.,
RA Iizuka H., Mishiro S., Miyakawa Y., Mayumi M.;
RT "Hepatitis C virus variants from Vietnam are classifiable into the
RT seventh, eighth, and ninth major genetic groups.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:11022-11026 (1994).

[2]
RP SEQUENCE FROM N.A.
RA Okamoto H.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; D88468; BAA13617.1; -.
DR PIR; PQ0804; PQ0804.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; P:structural molecule activity; IEA.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 414
SQ SEQUENCE 414 AA; 44937 MW; 65355640863B3DCF CRC64;

Query Match 44.5%; Score 85; DB 2; Length 414;
Best Local Similarity 100.0%; Pred. No. 8.4e-75;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 WGNPDRRRNLGKVIDITLCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVAT 166
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 167 GNLPCCSFIFLLALLSCLITTPASA 191
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 14
Q82531
ID Q82531 PRELIMINARY; PRT; 3016 AA.
AC Q82531;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98378034; PubMed=9714232;
RA Tokita H., Okamoto H., Iizuka H., Kishimoto J., Taada F., Miyakawa Y.,
RA Mayumi M.;
RT "The entire nucleotide sequences of three hepatitis C virus isolates
RT in genetic groups 7-9 and comparison with those in the other eight
RT genetic groups.";
RL J. Gen. Virol. 79:1847-1857 (1998).
DR EMBL; D84264; BAA32666.1; -.
DR PIR; PQ0804; PQ0804.
DR HSSP; P27958; 1A1V.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; P:structural molecule activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
```

DR GO:0019079; P:vital genome replication; IEA.  
DR GO:0019087; P:vital transformation; IEA.  
DR InterPro; IPR00345; Cyt\_c\_heme\_BS.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR002522; HCV capsid.  
DR InterPro; IPR002521; HCV core.  
DR InterPro; IPR002531; HCV NS1.  
DR InterPro; IPR002531; HCV NS1.  
DR InterPro; IPR000745; HCV NS4a.  
DR InterPro; IPR001490; HCV NS4b.  
DR InterPro; IPR002868; HCV NS5a.  
DR InterPro; IPR002166; HCV RdRP.  
DR InterPro; IPR004109; Peptidase S29.  
DR InterPro; IPR009003; Pept Ser Cys.  
DR InterPro; IPR002518; Pept\_U39\_HCV\_NS2.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR007094; RNA\_pol\_PSVir.  
DR Pfam; PF01543; HCV capsid; 1.  
DR Pfam; PF01542; HCV core; 1.  
DR Pfam; PF01539; HCV env; 1.  
DR Pfam; PF01560; HCV NS1; 1.  
DR Pfam; PF01538; HCV NS1; 1.  
DR Pfam; PF02907; HCV NS2; 1.  
DR Pfam; PF01006; HCV NS4a; 1.  
DR Pfam; PF01001; HCV NS4b; 1.  
DR Pfam; PF01506; HCV NS5a; 1.  
DR Pfam; PF00998; Viral\_RGRP; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN 1.  
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
KW Polyprotein; Transmembrane.  
SQ SEQUENCE 3016 AA; 328032 MW; 4E5CFF96258BCE3B CRC64;

Query Match 44.5%; Score 85; DB 2; Length 3016;  
Best Local Similarity 100.0%; Pred. No. 4.1e-74;  
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 WGNPDRRRNRNLGKVIDTLTCGPADLMGYPVVGAPLGVAALAHGVRAIEDGINYAT 166  
DB 107 WGNPDRRRNRNLGKVIDTLTCGPADLMGYPVVGAPLGVAALAHGVRAIEDGINYAT 166

QY 167 GNLPGCSFISFLALLSLCTTPASA 191  
DB 167 GNLPGCSFISFLALLSLCTTPASA 191

RESULT 15  
Q81271 ID Q81271 PRELIMINARY; PRT; 414 AA.  
AC Q81271;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Core, env and part of E2/NS1 (Fragment).  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96005057; PubMed=7561773;  
RA Tokita H., Okamoto H., Luengrojjanakul P., Varesangthip K.,  
RA Chainuvati T., Lizuka H., Tsuda F., Miyakawa Y., Mayumi M.;  
RT "Hepatitis C virus variants from Thailand classifiable into five novel  
RT genotypes in the sixth (6b), seventh (7c,7d) and ninth (9b, 9c) major  
RT genetic groups.";  
RL J. Gen. Virol. 76:2329-2335(1995).  
DR EMBL; D37846; BAA07092.1;  
DR GO:0016021; C:integral to membrane; IEA.  
DR GO:0019028; C:viral capsid; IEA.  
DR GO:0019031; C:viral envelope; IEA.  
DR GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR002522; HCV capsid.

DR InterPro; IPR002521; HCV core.  
DR InterPro; IPR002519; HCV env.  
DR InterPro; IPR002531; HCV NS1.  
DR Pfam; PF01543; HCV capsid; 1.  
DR Pfam; PF01542; HCV core; 1.  
DR Pfam; PF01539; HCV env; 1.  
DR Pfam; PF01560; HCV NS1; 1.  
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
KW Polyprotein; Transmembrane.  
FT NON\_TER 414 414  
SQ SEQUENCE 414 AA; 45148 MW; D63EE7CED5B71776 CRC64;

Query Match 42.9%; Score 82; DB 2; Length 414;  
Best Local Similarity 100.0%; Pred. No. 7.5e-72;  
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 WAQPGYPWPLYGNEGCGWAGWLLSPRGSRPHWGPNDPRRSRLGKVIDTLTCGFADLMG 135  
DB 76 WAQPGYPWPLYGNEGCGWAGWLLSPRGSRPHWGPNDPRRSRLGKVIDTLTCGFADLMG 135

QY 136 YIPVVGAPLGVAALAHGVRA 157  
DB 136 YIPVVGAPLGVAALAHGVRA 157

Search completed: October 30, 2004, 01:59:29  
Job time : 68 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2004, 01:57:19 ; Search time 21 Seconds  
(without alignments)  
603.178 Million cell updates/sec

Title: US-09-084-691B-206

Perfect score: 191

Sequence: 1 MSTLPKQKTKRNTNRPT.....CSFSIFLLALLSLITTPASA 191

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 478139 seqs, 66318000 residues

Word size : 0

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2.6/prodata/1/1aa/5A COMB.pep:\*
- 2: /cgn2.6/prodata/1/1aa/5B COMB.pep:\*
- 3: /cgn2.6/prodata/1/1aa/6A COMB.pep:\*
- 4: /cgn2.6/prodata/1/1aa/6B COMB.pep:\*
- 5: /cgn2.6/prodata/1/1aa/PCTUS COMB.pep:\*
- 6: /cgn2.6/prodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	191	100.0	191	2	US-08-290-665A-206
2	191	100.0	191	5	PCT-US95-10398-206
3	191	100.0	319	4	US-08-635-886C-228
4	191	100.0	319	4	US-08-974-690C-228
5	51	26.7	191	2	US-08-290-665A-161
6	51	26.7	191	2	US-08-290-665A-167
7	51	26.7	191	2	US-08-290-665A-168
8	51	26.7	191	2	US-08-290-665A-172
9	51	26.7	191	2	US-08-290-665A-174
10	51	26.7	191	2	US-08-290-665A-191
11	51	26.7	191	2	US-08-290-665A-198
12	51	26.7	191	2	US-08-290-665A-199
13	51	26.7	191	2	US-08-290-665A-200
14	51	26.7	191	2	US-08-290-665A-201
15	51	26.7	191	2	US-08-290-665A-202
16	51	26.7	191	2	US-08-290-665A-203
17	51	26.7	191	2	US-08-290-665A-205
18	51	26.7	191	5	PCT-US95-10398-161
19	51	26.7	191	5	PCT-US95-10398-167
20	51	26.7	191	5	PCT-US95-10398-168
21	51	26.7	191	5	PCT-US95-10398-172
22	51	26.7	191	5	PCT-US95-10398-174
23	51	26.7	191	5	PCT-US95-10398-191
24	51	26.7	191	5	PCT-US95-10398-198
25	51	26.7	191	5	PCT-US95-10398-199
26	51	26.7	191	5	PCT-US95-10398-200
27	51	26.7	191	5	PCT-US95-10398-201

28	51	26.7	191	5	PCT-US95-10398-202	Sequence 202, App
29	51	26.7	191	5	PCT-US95-10398-203	Sequence 203, App
30	51	26.7	191	5	PCT-US95-10398-205	Sequence 205, App
31	51	26.7	450	4	US-08-635-886C-189	Sequence 189, App
32	51	26.7	450	4	US-08-635-886C-194	Sequence 194, App
33	51	26.7	450	4	US-08-974-690C-189	Sequence 189, App
34	51	26.7	450	4	US-08-974-690C-194	Sequence 194, App
35	51	26.7	3010	4	US-09-539-601-3	Sequence 3, Appli
36	51	26.7	3010	4	US-09-539-601-21	Sequence 21, Appl
37	51	26.7	3010	4	US-09-539-601-27	Sequence 27, Appl
38	51	26.7	3010	4	US-09-539-601-33	Sequence 33, Appl
39	50	26.2	191	2	US-08-290-665A-197	Sequence 197, App
40	50	26.2	191	5	PCT-US95-10398-197	Sequence 197, App
41	50	26.2	319	4	US-08-635-886C-217	Sequence 217, App
42	50	26.2	319	4	US-08-974-690C-217	Sequence 217, App
43	49	25.7	75	4	US-08-905-054B-1	Sequence 1, Appli
44	49	25.7	75	5	PCT-US92-07813-1	Sequence 1, Appli
45	49	25.7	90	1	US-07-681-703B-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1  
US-08-290-665A-206  
; Sequence 206, Application US/08290665A  
; Patent No. 5882852  
; GENERAL INFORMATION:  
; APPLICANT: BUKH, J., MILLER, R.H. AND  
; APPLICANT: PURCELL, R.H.  
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND  
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS  
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE  
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES  
; NUMBER OF SEQUENCES: 263  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/290,665A  
; FILING DATE: 15-AUG-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: RICHARD W. BORK  
; REGISTRATION NUMBER: 36,459  
; REFERENCE/DOCKET NUMBER: 2026-4116  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 758-4800  
; TELEFAX: (212) 751-6849  
; TELEX: 421792  
; INFORMATION FOR SEQ ID NO: 206:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 191 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; ORIGINAL SOURCE:  
; ORGANISM: homosapiens  
; INDIVIDUAL ISOLATE: HK2  
; US-08-290-665A-206

Query Match 100.0%; Score 191; DB 2; Length 191;  
Best Local Similarity 100.0%; Pred. No. 1e-173;

Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSTLPKPKQKTKNTNRPTDVKFPGGQIVGGVYLLPRGPRLGVRAATKTSERSQPRG 60  
 Db 1 MSTLPKPKQKTKNTNRPTDVKFPGGQIVGGVYLLPRGPRLGVRAATKTSERSQPRG 60

QY 61 RRQIPKARQPGQRHWAQPGYPWPLYGNECGWAGWLLSPRGRPHWGPNDPRRSRNLG 120  
 Db 61 RRQIPKARQPGQRHWAQPGYPWPLYGNECGWAGWLLSPRGRPHWGPNDPRRSRNLG 120

QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
 Db 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191  
 Db 181 LLSCLTTPASA 191

RESULT 2  
 PCT-US95-10398-206  
 ; Sequence 206, Application PC/TUS9510398  
 ; GENERAL INFORMATION:  
 ; APPLICANT: BUKH, J., MILLER, R.H. AND  
 ; APPLICANT: PURCELL, R.H.  
 ; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
 ; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND  
 ; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS  
 ; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE  
 ; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES  
 ; NUMBER OF SEQUENCES: 263  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MORGAN & FINNEGAN  
 ; STREET: 345 PARK AVENUE  
 ; CITY: NEW YORK  
 ; STATE: NEW YORK  
 ; COUNTRY: USA  
 ; ZIP: 10154  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: FLOPPY DISK  
 ; COMPUTER: IBM PC COMPATIBLE  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: WORDPERFECT 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US95/10398  
 ; FILING DATE: 15-AUG-1995  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/086,428  
 ; FILING DATE: 29 JUNE 1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/290/665  
 ; FILING DATE: 15 AUGUST 1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: RICHARD W. BORK  
 ; REGISTRATION NUMBER: 36,459  
 ; REFERENCE/DOCKET NUMBER: 2026-4116  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 758-4800  
 ; TELEFAX: (212) 751-6849  
 ; TELEX: 421792  
 ; INFORMATION FOR SEQ ID NO: 206:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 191 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: unknown  
 ; TOPOLOGY: unknown  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: homopapiens  
 ; INDIVIDUAL ISOLATE: HK2  
 ; PCT-US95-10398-206

Query Match 100.0%; Score 191; DB 5; Length 191;

Best Local Similarity 100.0%; Pred. No. 1e-173;  
 Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSTLPKPKQKTKNTNRPTDVKFPGGQIVGGVYLLPRGPRLGVRAATKTSERSQPRG 60  
 Db 1 MSTLPKPKQKTKNTNRPTDVKFPGGQIVGGVYLLPRGPRLGVRAATKTSERSQPRG 60

QY 61 RRQIPKARQPGQRHWAQPGYPWPLYGNECGWAGWLLSPRGRPHWGPNDPRRSRNLG 120  
 Db 61 RRQIPKARQPGQRHWAQPGYPWPLYGNECGWAGWLLSPRGRPHWGPNDPRRSRNLG 120

QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
 Db 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191  
 Db 181 LLSCLTTPASA 191

RESULT 3  
 US-08-635-886C-228  
 ; Sequence 228, Application US/08635886C  
 ; Patent No. 6555114  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LEROUX-ROELS, Geert  
 ; APPLICANT: DELEYS, Robert  
 ; APPLICANT: MAERTENS, Geert  
 ; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C  
 ; TITLE OF INVENTION: VIRUS  
 ; FILE REFERENCE: 2752-18  
 ; CURRENT APPLICATION NUMBER: US/08/635,886C  
 ; CURRENT FILING DATE: 1996-04-25  
 ; PRIOR APPLICATION NUMBER: PCT/EP94/03555  
 ; PRIOR FILING DATE: 1994-10-28  
 ; PRIOR APPLICATION NUMBER: EP 93402718.6  
 ; PRIOR FILING DATE: 1993-11-04  
 ; NUMBER OF SEQ ID NOS: 286  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 228  
 ; LENGTH: 319  
 ; TYPE: PRT  
 ; ORGANISM: hepatitis C virus  
 ; US-08-635-886C-228

Query Match 100.0%; Score 191; DB 4; Length 319;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-173;  
 Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSTLPKPKQKTKNTNRPTDVKFPGGQIVGGVYLLPRGPRLGVRAATKTSERSQPRG 60  
 Db 1 MSTLPKPKQKTKNTNRPTDVKFPGGQIVGGVYLLPRGPRLGVRAATKTSERSQPRG 60

QY 61 RRQIPKARQPGQRHWAQPGYPWPLYGNECGWAGWLLSPRGRPHWGPNDPRRSRNLG 120  
 Db 61 RRQIPKARQPGQRHWAQPGYPWPLYGNECGWAGWLLSPRGRPHWGPNDPRRSRNLG 120

QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
 Db 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180

QY 181 LLSCLTTPASA 191  
 Db 181 LLSCLTTPASA 191

RESULT 4  
 US-08-974-690C-228  
 ; Sequence 228, Application US/08974690C  
 ; Patent No. 6613333  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LEROUX-ROELS, Geert  
 ; APPLICANT: DELEYS, Robert

```

; APPLICANT: MAERTENS, Geert
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C
; TITLE OF INVENTION: VIRUS
; FILE REFERENCE: 2551-94
; CURRENT APPLICATION NUMBER: US/08/974,690C
; CURRENT FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: PCT/EP94/03555
; PRIOR FILING DATE: 1994-10-28
; PRIOR APPLICATION NUMBER: EP 93402718.6
; PRIOR FILING DATE: 1993-11-04
; NUMBER OF SEQ ID NOS: 286
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 228
; LENGTH: 319
; TYPE: PRT
; ORGANISM: hepatitis C virus
; US-08-974-690C-228

Query Match      100.0%; Score 191; DB 4; Length 319;
Best Local Similarity 100.0%; Pred. No. 1.6e-173;
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  MSTLPKQRTKNTNRRPTDVKPPGGQIVGGVYLLPRRGPRLGVATRKTSRSQPRG 60
Db      1  MSTLPKQRTKNTNRRPTDVKPPGGQIVGGVYLLPRRGPRLGVATRKTSRSQPRG 60

QY      61  RRQIPKAROQGRHWAQGYPPWPLYNCGCGWAGWLLSPGSRPHWGPNDPRRSNLTG 120
Db      61  RRQIPKAROQGRHWAQGYPPWPLYNCGCGWAGWLLSPGSRPHWGPNDPRRSNLTG 120

QY      121  KVTDLTTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINATGNLPGCCSFSIFLLA 180
Db      121  KVTDLTTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINATGNLPGCCSFSIFLLA 180

QY      181  LLSCLATTPASA 191
Db      181  LLSCLATTPASA 191

```

```

1  RESULT 5
2  US-08-290-665A-161
3  ; Sequence 161, Application US/08290665A
4  ; Patent No. 5882852
5  ; GENERAL INFORMATION:
6  ;
7  ; APPLICANT: BUKH, J., MILLER, R.H. A
8  ; APPLICANT: PORCELL, R.H.
9  ; TITLE OF INVENTION: NUCLEOTIDE AND
10 ; TITLE OF INVENTION: AMINO ACID SEQ
11 ; TITLE OF INVENTION: CORE GENES OF J
12 ; TITLE OF INVENTION: AND THE USE OF I
13 ; TITLE OF INVENTION: SEQUENCES IN D
14 ; NUMBER OF SEQUENCES: 263
15 ; CORRESPONDENCE ADDRESS:
16 ; ADDRESSEE: MORGAN & FINNEGAN
17 ; STREET: 345 PARK AVENUE
18 ; CITY: NEW YORK
19 ; STATE: NEW YORK
20 ; COUNTRY: USA
21 ; ZIP: 10154
22 ; COMPUTER READABLE FORM:
23 ; MEDIUM TYPE: FLOPPY DISK
24 ; COMPUTER: IBM PC COMPATIBLE
25 ; OPERATING SYSTEM: PC-DOS/MS-DOS
26 ; SOFTWARE: WORDPERFECT 5.1
27 ; CURRENT APPLICATION DATA:
28 ; APPLICATION NUMBER: US/08/290,665A
29 ; FILING DATE: 15-AUG-1994
30 ; CLASSIFICATION: 435
31 ; ATTORNEY/AGENT INFORMATION:
32 ; NAME: RICHARD W. BORK
33 ; REGISTRATION NUMBER: 36,459
34 ; REFERENCE/DOCKET NUMBER: 2026-411
35 ; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 161:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: SA10
US-08-290-665A-161

Query Match          26.7%; Score 51; DB 2; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.1e-40;
Matches 51; Conservative 0; Mismatches 0; Indels 0;

Qy 21 DVKFFGGGQIVGGVYLLPRRGRPLGVATRKTSERSQPRGRQPIPKARQP 71
    |||
Db 21 DVKFFGGGQIVGGVYLLPRRGRPLGVATRKTSERSQPRGRQPIPKARQP 71

RESULT 6
US-08-290-665A-167
; Sequence 167, Application US/08290665A
; Patent No. 5882852
; GENERAL INFORMATION:
; APPLICANT: BUKH, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,665A
; FILING DATE: 15-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 167:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: T10
US-08-290-665A-167

Query Match          26.7%; Score 51; DB 2; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.1e-40;

```

Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKPPGGQIVGGVYLLPRGPRGLGVRATKTSRSQPRGRQPIPKARQP 71  
|||||  
Db 21 DVKPPGGQIVGGVYLLPRGPRGLGVRATKTSRSQPRGRQPIPKARQP 71  
|||||

## RESULT 7

US-08-290-665A-168  
; Sequence 168, Application US/08290665A  
; Patent No. 5882852  
; GENERAL INFORMATION:  
; APPLICANT: BUKH, J., MILLER, R.H. AND  
; APPLICANT: PURCELL, R.H.  
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND  
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS  
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE  
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES  
; NUMBER OF SEQUENCES: 263  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/290,665A  
; FILING DATE: 15-AUG-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: RICHARD W. BORK  
; REGISTRATION NUMBER: 36,459  
; REFERENCE/DOCKET NUMBER: 2026-4116  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 758-4800  
; TELEFAX: (212) 751-6849  
; TELEX: 421792  
; INFORMATION FOR SEQ ID NO: 168:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 191 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; ORGANISM: homosapiens  
; INDIVIDUAL ISOLATE: SW2  
US-08-290-665A-168

Query Match 26.7%; Score 51; DB 2; Length 191;  
Best Local Similarity 100.0%; Pred. No. 1.1e-40;  
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKPPGGQIVGGVYLLPRGPRGLGVRATKTSRSQPRGRQPIPKARQP 71  
|||||  
Db 21 DVKPPGGQIVGGVYLLPRGPRGLGVRATKTSRSQPRGRQPIPKARQP 71  
|||||

## RESULT 8

US-08-290-665A-172  
; Sequence 172, Application US/08290665A  
; Patent No. 5882852  
; GENERAL INFORMATION:  
; APPLICANT: BUKH, J., MILLER, R.H. AND  
; APPLICANT: PURCELL, R.H.  
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND

; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS  
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE  
; NUMBER OF SEQUENCES: 263  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/290,665A  
; FILING DATE: 15-AUG-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: RICHARD W. BORK  
; REGISTRATION NUMBER: 36,459  
; REFERENCE/DOCKET NUMBER: 2026-4116  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 758-4800  
; TELEFAX: (212) 751-6849  
; TELEX: 421792  
; INFORMATION FOR SEQ ID NO: 172:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 191 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; ORIGINAL SOURCE:  
; ORGANISM: homosapiens  
; INDIVIDUAL ISOLATE: HK3  
US-08-290-665A-172

Query Match 26.7%; Score 51; DB 2; Length 191;  
Best Local Similarity 100.0%; Pred. No. 1.1e-40;  
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKPPGGQIVGGVYLLPRGPRGLGVRATKTSRSQPRGRQPIPKARQP 71  
|||||  
Db 21 DVKPPGGQIVGGVYLLPRGPRGLGVRATKTSRSQPRGRQPIPKARQP 71  
|||||

## RESULT 9

US-08-290-665A-174  
; Sequence 174, Application US/08290665A  
; Patent No. 5882852  
; GENERAL INFORMATION:  
; APPLICANT: BUKH, J., MILLER, R.H. AND  
; APPLICANT: PURCELL, R.H.  
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND  
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS  
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE  
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES  
; NUMBER OF SEQUENCES: 263  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORDPERFECT 5.1



RESULT 12  
US-08-290-665A-199  
; Sequence 199, Application US/08290665A  
; Patent No. 5882852  
; GENERAL INFORMATION:  
; APPLICANT: BURKH, J., MILLER, R.H. AND  
; APPLICANT: PURCELL, R.H.  
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND  
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS  
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE  
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES  
; NUMBER OF SEQUENCES: 263  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/290,665A  
; FILING DATE: 15-AUG-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: RICHARD W. BORK  
; REGISTRATION NUMBER: 36,459  
; REFERENCE/DOCKET NUMBER: 2026-4116  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 758-4800  
; TELEFAX: (212) 751-6849  
; TELEX: 421792  
; INFORMATION FOR SEQ ID NO: 199:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 191 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; ORGANISM: homosapiens  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: SA5  
; US-08-290-665A-199  
  
Query Match 26.7%; Score 51; DB 2; Length 191;  
Best Local Similarity 100.0%; Pred. No. 1.1e-40;  
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 21 DVKFPGGGQIVGGVYLLPRGPRGLGVRATKTSERSQPRGRQPIPKARQP 71  
Db 21 DVKFPGGGQIVGGVYLLPRGPRGLGVRATKTSERSQPRGRQPIPKARQP 71  
  
RESULT 13  
US-08-290-665A-200  
; Sequence 200, Application US/08290665A  
; Patent No. 5882852  
; GENERAL INFORMATION:  
; APPLICANT: BURKH, J., MILLER, R.H. AND  
; APPLICANT: PURCELL, R.H.  
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND  
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS  
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE  
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES  
; NUMBER OF SEQUENCES: 263  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE

; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/290,665A  
; FILING DATE: 15-AUG-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: RICHARD W. BORK  
; REGISTRATION NUMBER: 36,459  
; REFERENCE/DOCKET NUMBER: 2026-4116  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 758-4800  
; TELEFAX: (212) 751-6849  
; TELEX: 421792  
; INFORMATION FOR SEQ ID NO: 200:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 191 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; ORGANISM: homosapiens  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: SA7  
; US-08-290-665A-200  
  
Query Match 26.7%; Score 51; DB 2; Length 191;  
Best Local Similarity 100.0%; Pred. No. 1.1e-40;  
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 21 DVKFPGGGQIVGGVYLLPRGPRGLGVRATKTSERSQPRGRQPIPKARQP 71  
Db 21 DVKFPGGGQIVGGVYLLPRGPRGLGVRATKTSERSQPRGRQPIPKARQP 71  
  
RESULT 14  
US-08-290-665A-201  
; Sequence 201, Application US/08290665A  
; Patent No. 5882852  
; GENERAL INFORMATION:  
; APPLICANT: BURKH, J., MILLER, R.H. AND  
; APPLICANT: PURCELL, R.H.  
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND  
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS  
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE  
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES  
; NUMBER OF SEQUENCES: 263  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/290,665A  
; FILING DATE: 15-AUG-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: RICHARD W. BORK  
; REGISTRATION NUMBER: 36,459



REFERENCE/DOCKET NUMBER: 2026-4116  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 201:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 191 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
ORIGINAL SOURCE:  
ORGANISM: homosapiens  
INDIVIDUAL ISOLATE: SA1  
US-08-290-665A-201

Query Match 26.7%; Score 51; DB 2; Length 191;  
Best Local Similarity 100.0%; Pred. No. 1.1e-40;  
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKFPGGGQIVGGVILLPRRGLGVRAIRKTSERSQPRGRQPIPKARQP 71  
DB 21 DVKFPGGGQIVGGVILLPRRGLGVRAIRKTSERSQPRGRQPIPKARQP 71

RESULT 15  
US-08-290-665A-202  
Sequence 202, Application US/08290665A  
Patent No. 5882852  
GENERAL INFORMATION:  
APPLICANT: BUKH, J., MILLER, R.H. AND  
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND  
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS  
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE  
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES  
NUMBER OF SEQUENCES: 263  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/290,665A  
FILING DATE: 15-AUG-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: RICHARD W. BORK  
REGISTRATION NUMBER: 36,459  
REFERENCE/DOCKET NUMBER: 2026-4116  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 202:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 191 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
ORIGINAL SOURCE:  
ORGANISM: homosapiens  
INDIVIDUAL ISOLATE: SA3  
US-08-290-665A-202

Query Match 26.7%; Score 51; DB 2; Length 191;  
Best Local Similarity 100.0%; Pred. No. 1.1e-40;  
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKFPGGGQIVGGVILLPRRGLGVRAIRKTSERSQPRGRQPIPKARQP 71  
DB 21 DVKFPGGGQIVGGVILLPRRGLGVRAIRKTSERSQPRGRQPIPKARQP 71

Search completed: October 30, 2004, 02:00:42  
Job time : 21 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 30, 2004, 01:58:21 ; Search time 49 Seconds

(without alignments)  
1263.780 Million cell updates/sec

Title: US-09-084-691B-206

Perfect score: 191  
Sequence: 1 MSTLPKPKQKTKRNTNRPT.....CSFSIFLLALLSCLTPASA 191

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1370721 seqs, 324215800 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1370721

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pcp:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pcp:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US05\_NEW\_PUB.pcp:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pcp:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pcp:\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pcp:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pcp:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pcp:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pcp:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pcp:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pcp:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pcp:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pcp:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pcp:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pcp:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pcp:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pcp:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pcp:\*  
19: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pcp:\*  
20: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pcp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	191	100.0	319	15	US-10-651-165-228
2	51	26.7	450	15	US-10-651-165-189
3	51	26.7	450	15	US-10-651-165-194
4	51	26.7	3010	15	US-10-467-000-1
5	50	26.2	319	15	US-10-651-165-217
6	49	25.7	75	15	US-10-431-587-1
7	49	25.7	77	9	US-09-921-397-3
8	49	25.7	91	9	US-09-758-308-1
9	49	25.7	94	10	US-09-891-983A-16
10	49	25.7	94	14	US-10-173-480-16
11	49	25.7	94	16	US-10-753-910-16
12	49	25.7	96	10	US-09-899-046-192
13	49	25.7	96	10	US-09-878-281-192

14	49	25.7	96	10	US-09-873-224-192
15	49	25.7	97	9	US-09-756-875-8
16	49	25.7	103	9	US-09-921-397-77
17	49	25.7	113	9	US-09-921-397-78
18	49	25.7	117	9	US-09-851-138-28
19	49	25.7	117	15	US-10-651-165-225
20	49	25.7	120	9	US-09-306-780-4
21	49	25.7	122	14	US-10-098-857B-1
22	49	25.7	126	10	US-09-899-046-166
23	49	25.7	126	10	US-09-878-281-166
24	49	25.7	126	10	US-09-873-224-166
25	49	25.7	130	14	US-10-268-569-19
26	49	25.7	150	9	US-09-306-780-6
27	49	25.7	151	14	US-10-292-129-14
28	49	25.7	158	9	US-09-851-138-66
29	49	25.7	161	9	US-09-306-780-8
30	49	25.7	166	10	US-09-899-046-164
31	49	25.7	166	10	US-09-899-046-194
32	49	25.7	166	10	US-09-878-281-164
33	49	25.7	166	10	US-09-878-281-194
34	49	25.7	166	10	US-09-873-224-164
35	49	25.7	166	10	US-09-873-224-194
36	49	25.7	176	9	US-09-306-780-20
37	49	25.7	182	9	US-09-929-955-2
38	49	25.7	182	13	US-10-104-966-2
39	49	25.7	182	15	US-10-719-619-2
40	49	25.7	190	14	US-10-268-562-1
41	49	25.7	190	15	US-10-450-649-7
42	49	25.7	191	9	US-09-306-780-10
43	49	25.7	191	10	US-09-194-949-3
44	49	25.7	235	15	US-10-365-620-58
45	49	25.7	249	15	US-10-365-620-54

#### ALIGNMENTS

#### RESULT 1

US-10-651-165-228  
; Sequence 228, Application US/10651165  
; Publication No. US20040047877A1  
; GENERAL INFORMATION:  
; APPLICANT: LEROUX-ROELS, Geert  
; APPLICANT: DELEUX, Robert  
; APPLICANT: MAERTENS, Geert  
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C VIRUS  
; FILE REFERENCE: 2551-94  
; CURRENT APPLICATION NUMBER: US/10/651,165  
; PRIOR FILING DATE: 2003-09-02  
; PRIOR APPLICATION NUMBER: US/08/974,690C  
; PRIOR FILING DATE: 1997-11-19  
; PRIOR APPLICATION NUMBER: PCT/EP94/03555  
; PRIOR FILING DATE: 1994-10-28  
; PRIOR APPLICATION NUMBER: EP 93402718.6  
; PRIOR FILING DATE: 1993-11-04  
; NUMBER OF SEQ ID NOS: 286  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 228  
; LENGTH: 319  
; TYPE: PPT  
; ORGANISM: hepatitis C virus  
US-10-651-165-228

Query Match 100.0%; Score 191; DB 15; Length 319;  
Best Local Similarity 100.0%; Pred. No. 6.5e-166;  
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSTLPKPKQKTKRNTNRPTDKPFGGQIVGGVYLLPRRGPRLGVRATKTSRSQPRG 60

Db 1 MSTLPKPKQKTKRNTNRPTDKPFGGQIVGGVYLLPRRGPRLGVRATKTSRSQPRG 60

QY 61 RQPIPKARQPKRHWAPGYVFWPLYGNEGCGWAGLLSPRGRPHWGPNDPRRSRLG 120

Db 61 RQPIPKARQPGQHWAPGYPPLVYNGECCHAGWLLSPRGSRPHWGNDRPRSRNLG 120  
QY 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
Db 121 KVIDTLTCGFADLMGYIPVVGAPLGGVAAALAHGVRAIEDGINVATGNLPGCSFSIFLLA 180  
QY 181 LLSCLTTPASA 191  
Db 181 LLSCLTTPASA 191

RESULT 2  
US-10-651-165-189  
; Sequence 189, Application US/10651165  
; Publication No. US20040047877A1  
; GENERAL INFORMATION:  
; APPLICANT: LEROUX-ROELS, Geert  
; APPLICANT: DELEYS, Robert  
; APPLICANT: MAERTENS, Geert  
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C  
; TITLE OF INVENTION: VIRUS  
; FILE REFERENCE: 2551-94  
; CURRENT APPLICATION NUMBER: US/10/651.165  
; CURRENT FILING DATE: 2003-09-02  
; PRIOR APPLICATION NUMBER: US/08/974,690C  
; PRIOR FILING DATE: 1997-11-19  
; PRIOR APPLICATION NUMBER: PCT/EP94/03555  
; PRIOR FILING DATE: 1994-10-28  
; PRIOR APPLICATION NUMBER: EP 93402718.6  
; PRIOR FILING DATE: 1993-11-04  
; NUMBER OF SEQ ID NOS: 286  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 189  
; LENGTH: 450  
; TYPE: PRT  
; ORGANISM: hepatitis C virus  
US-10-651-165-189

Query Match 26.7%; Score 51; DB 15; Length 450;  
Best Local Similarity 100.0%; Pred. No. 5.1e-38;  
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKFPGGQIVGGVYLLPRRGPRLGVRATKTSERSQPRGRQPIPKARQP 71  
Db 21 DVKFPGGQIVGGVYLLPRRGPRLGVRATKTSERSQPRGRQPIPKARQP 71

RESULT 3  
US-10-651-165-194  
; Sequence 194, Application US/10651165  
; Publication No. US20040047877A1  
; GENERAL INFORMATION:  
; APPLICANT: LEROUX-ROELS, Geert  
; APPLICANT: DELEYS, Robert  
; APPLICANT: MAERTENS, Geert  
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C  
; TITLE OF INVENTION: VIRUS  
; FILE REFERENCE: 2551-94  
; CURRENT APPLICATION NUMBER: US/10/651.165  
; CURRENT FILING DATE: 2003-09-02  
; PRIOR APPLICATION NUMBER: US/08/974,690C  
; PRIOR FILING DATE: 1997-11-19  
; PRIOR APPLICATION NUMBER: PCT/EP94/03555  
; PRIOR FILING DATE: 1994-10-28  
; PRIOR APPLICATION NUMBER: EP 93402718.6  
; PRIOR FILING DATE: 1993-11-04  
; NUMBER OF SEQ ID NOS: 286  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 194  
; LENGTH: 450  
; TYPE: PRT  
; ORGANISM: hepatitis C virus

US-10-651-165-194  
Query Match 26.7%; Score 51; DB 15; Length 450;  
Best Local Similarity 100.0%; Pred. No. 5.1e-38;  
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKFPGGQIVGGVYLLPRRGPRLGVRATKTSERSQPRGRQPIPKARQP 71  
Db 21 DVKFPGGQIVGGVYLLPRRGPRLGVRATKTSERSQPRGRQPIPKARQP 71

RESULT 4  
US-10-467-000-1  
; Sequence 1, Application US/10467000  
; Publication No. US20040067486A1  
; GENERAL INFORMATION:  
; APPLICANT: De Francesco, Raffaele  
; APPLICANT: Migliaccio, Giovanni  
; APPLICANT: Paonessa, Giacomo  
; TITLE OF INVENTION: HEPATITIS C VIRUS REPLICONS AND REPLICON  
; TITLE OF INVENTION: ENHANCED CELLS  
; FILE REFERENCE: ITR0003P  
; CURRENT APPLICATION NUMBER: US/10/467,000  
; CURRENT FILING DATE: 2003-07-21  
; PRIOR APPLICATION NUMBER: PCT/EP02/00526  
; PRIOR FILING DATE: 2002-01-16  
; PRIOR APPLICATION NUMBER: 60/263,479  
; PRIOR FILING DATE: 2001-01-23  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 3010  
; TYPE: PRT  
; ORGANISM: Con 1 HCV isolate nucleic acid  
US-10-467-000-1

Query Match 26.7%; Score 51; DB 15; Length 3010;  
Best Local Similarity 100.0%; Pred. No. 2.5e-37;  
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKFPGGQIVGGVYLLPRRGPRLGVRATKTSERSQPRGRQPIPKARQP 71  
Db 21 DVKFPGGQIVGGVYLLPRRGPRLGVRATKTSERSQPRGRQPIPKARQP 71

RESULT 5  
US-10-651-165-217  
; Sequence 217, Application US/10651165  
; Publication No. US20040047877A1  
; GENERAL INFORMATION:  
; APPLICANT: LEROUX-ROELS, Geert  
; APPLICANT: DELEYS, Robert  
; APPLICANT: MAERTENS, Geert  
; TITLE OF INVENTION: IMMUNODOMINANT HUMAN T CELL EPITOPES OF HEPATITIS C  
; TITLE OF INVENTION: VIRUS  
; FILE REFERENCE: 2551-94  
; CURRENT APPLICATION NUMBER: US/10/651.165  
; CURRENT FILING DATE: 2003-09-02  
; PRIOR APPLICATION NUMBER: US/08/974,690C  
; PRIOR FILING DATE: 1997-11-19  
; PRIOR APPLICATION NUMBER: PCT/EP94/03555  
; PRIOR FILING DATE: 1994-10-28  
; PRIOR APPLICATION NUMBER: EP 93402718.6  
; PRIOR FILING DATE: 1993-11-04  
; NUMBER OF SEQ ID NOS: 286  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 217  
; LENGTH: 319  
; TYPE: PRT  
; ORGANISM: hepatitis C virus  
US-10-651-165-217

Query Match 26.2%; Score 50; DB 15; Length 319;

Best Local Similarity 100.0%; Pred. No. 3.1e-37;  
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKPPGGQIVGGVYLLPRGRLGVRATRTKTSERSQPRGRQPIPKAR 70  
DB 21 DVKPPGGQIVGGVYLLPRGRLGVRATRTKTSERSQPRGRQPIPKAR 70

RESULT 6  
US-10-431-587-1  
; Sequence 1, Application US/10431587  
; Publication No. US20040072267A1  
; GENERAL INFORMATION:  
; APPLICANT: BIORAD PASTEUR  
; TITLE OF INVENTION: Method for simultaneously detecting an antigen of, and an antibody  
; FILE REFERENCE: BET 03P0456  
; CURRENT APPLICATION NUMBER: US/10/431,587  
; PRIOR FILING DATE: 2003-05-08  
; PRIOR APPLICATION NUMBER: FR 0205808  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 75  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
US-10-431-587-1

Query Match 25.7%; Score 49; DB 15; Length 75;  
Best Local Similarity 100.0%; Pred. No. 7.7e-37;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKPPGGQIVGGVYLLPRGRLGVRATRTKTSERSQPRGRQPIPKAR 69  
DB 21 DVKPPGGQIVGGVYLLPRGRLGVRATRTKTSERSQPRGRQPIPKAR 69

RESULT 7  
US-09-921-397-3  
; Sequence 3, Application US/09921397  
; Patent No. US20020151484A1  
; GENERAL INFORMATION:  
; APPLICANT: HYBRIGENICS  
; TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a  
; FILE REFERENCE: B4809A - JAZ  
; CURRENT APPLICATION NUMBER: US/09/921,397  
; PRIOR FILING DATE: 2001-08-02  
; PRIOR APPLICATION NUMBER: EP 00402225.7  
; NUMBER OF SEQ ID NOS: 156  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 77  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
US-09-921-397-3

Query Match 25.7%; Score 49; DB 9; Length 77;  
Best Local Similarity 100.0%; Pred. No. 7.9e-37;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKPPGGQIVGGVYLLPRGRLGVRATRTKTSERSQPRGRQPIPKAR 69  
DB 8 DVKPPGGQIVGGVYLLPRGRLGVRATRTKTSERSQPRGRQPIPKAR 56

RESULT 8  
US-09-758-308-1  
; Sequence 1, Application US/09758308  
; Patent No. US20020090607A1

; GENERAL INFORMATION:  
; APPLICANT: HOWARD A. FIELDS AND YURY E. KHUDYAKOV  
; TITLE OF INVENTION: ANTIGENIC EPITOPES AND MOSAIC POLYPEPTIDES OF HEPATITIS C VIRUS  
; FILE REFERENCE: 14114.0349U2  
; CURRENT APPLICATION NUMBER: US/09/758,308  
; CURRENT FILING DATE: 2001-01-10  
; PRIOR APPLICATION NUMBER: 60/092,339  
; PRIOR FILING DATE: 1999-07-10  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 91  
; TYPE: PRT  
; ORGANISM: Hepatitis C Virus  
US-09-758-308-1

Query Match 25.7%; Score 49; DB 9; Length 91;  
Best Local Similarity 100.0%; Pred. No. 9e-37;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKPPGGQIVGGVYLLPRGRLGVRATRTKTSERSQPRGRQPIPKAR 69  
DB 21 DVKPPGGQIVGGVYLLPRGRLGVRATRTKTSERSQPRGRQPIPKAR 69

RESULT 9  
US-09-891-983A-16  
; Sequence 16, Application US/09891983A  
; Publication No. US20030108858A1  
; GENERAL INFORMATION:  
; APPLICANT: Abbott Laboratories  
; APPLICANT: Shah, Dinesh O.  
; APPLICANT: Dawson, George A.  
; APPLICANT: Muerhoff, A. Scott  
; APPLICANT: Jiang, Lily  
; APPLICANT: Gutierrez, Robin A.  
; APPLICANT: Leary, Thomas P.  
; APPLICANT: Desai, Suresh  
; APPLICANT: Stewart, James L.  
; TITLE OF INVENTION: Methods For The simultaneous Detection  
; FILE REFERENCE: 6821.US.O1  
; CURRENT APPLICATION NUMBER: US/09/891,983A  
; CURRENT FILING DATE: 2001-08-26  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16  
; LENGTH: 94  
; TYPE: PRT  
; ORGANISM: Hepatitis C Virus  
US-09-891-983A-16

Query Match 25.7%; Score 49; DB 10; Length 94;  
Best Local Similarity 100.0%; Pred. No. 9.3e-37;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKPPGGQIVGGVYLLPRGRLGVRATRTKTSERSQPRGRQPIPKAR 69  
DB 15 DVKPPGGQIVGGVYLLPRGRLGVRATRTKTSERSQPRGRQPIPKAR 63

RESULT 10  
US-10-173-480-16  
; Sequence 16, Application US/10173480  
; Publication No. US20030152948A1  
; GENERAL INFORMATION:  
; APPLICANT: Abbott Laboratories  
; APPLICANT: Shah, Dinesh O.  
; APPLICANT: Dawson, George A.  
; APPLICANT: Muerhoff, A. Scott  
; APPLICANT: Jiang, Lily  
; APPLICANT: Gutierrez, Robin A.

```

; APPLICANT: Leary, Thomas P.
; APPLICANT: Desai, Suresh
; APPLICANT: Stewart, James L.
; TITLE OF INVENTION: METHODS FOR THE SIMULTANEOUS DETECTION
; TITLE OF INVENTION: OF HCV ANTIGENS AND HCV ANTIBODIES
; FILE REFERENCE: 6821.US.P1
; CURRENT APPLICATION NUMBER: US/10/173,480
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: 09/891,983
; PRIOR FILING DATE: 2001-06-26
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Variant - HCV-Core Recombinant
US-10-753-910-16

Query Match          25.7%; Score 49; DB 14; Length 94;
Best Local Similarity 100.0%; Pred.No. 9.3e-37;
Matches 49; Conservative 0; Mismatches 0; Indels 0

QY 21 DVKPPGGQIVGGVYLLPRRGPRLGVRATKTSERSQPRGRQPIPKAR 69
      |||||||
Db 15 DVKPPGGQIVGGVYLLPRRGPRLGVRATKTSERSQPRGRQPIPKAR 63

RESULT 11
US-10-753-910-16
; Sequence 16, Application US/10753910
; Publication No. US20040185436A1
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Shan, Dinesh O.
; APPLICANT: Dawson, Greoge A.
; APPLICANT: Muerhoff, A. Scott
; APPLICANT: Jiang, Lily
; APPLICANT: Gutierrez, Robin A.
; APPLICANT: Leary, Thomas P.
; APPLICANT: Desai, Suresh
; APPLICANT: Stewart, James L.
; TITLE OF INVENTION: METHODS FOR THE SIMULTANEOUS DETECTION
; TITLE OF INVENTION: OF HCV ANTIGENS AND HCV ANTIBODIES
; FILE REFERENCE: 6821.US.P1
; CURRENT APPLICATION NUMBER: US/10/753,910
; CURRENT FILING DATE: 2004-01-07
; PRIOR APPLICATION NUMBER: US/10/173,480
; PRIOR FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: 09/891,983
; PRIOR FILING DATE: 2001-06-26
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Variant - HCV-Core Recombinant
US-10-753-910-16

```

```

; Sequence 192, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 192:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 96 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-899-046-192

Query Match 25.7%; Score 49; DB 10; Length 96;
Best Local Similarity 100.0%; Pred. No. 9.5e-37;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKFPGGGQIVGGVILLPRRGPRGLGVRATKTSRSPQRRRQPIPKAR 69
Db 21 DVKFPGGGQIVGGVILLPRRGPRGLGVRATKTSRSPQRRRQPIPKAR 69

RESULT 13
US-09-878-281-192
; Sequence 192, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/878,281
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 192:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 96 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-878-281-192

```

RESULT 12  
US-09-899-046-192

```
US-09-873-224-192
; Sequence 192, Application US/09873224
; Publication No. US20030064360A1
; GENERAL INFORMATION:
; APPLICANT: <Unknown>
; TITLE OF INVENTION: New sequences of hepatitis C virus
; genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; CORRESPONDENCE ADDRESS:
; STREET: Industriepark Zwijnaarde 7, box 4
; CITY: Ghent
; COUNTRY: Belgium
; ZIP: B-9052
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/873,224
; FILING DATE: 05-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Innogenetics sa.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 00 32 9 241 07 11
; TELEFAX: 00 32 9 241 07 99
; INFORMATION FOR SEQ ID NO: 192:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 96 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 192:
US-09-873-224-192

Query Match 25.7%; Score 49; DB 10; Length 96;
Best Local Similarity 100.0%; Pred. No. 9.5e-37;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKFPGGQIVGGVLLPRRGRLGVRATRKTSERSQPRGRQPIPKAR 69
Db 21 DVKFPGGQIVGGVLLPRRGRLGVRATRKTSERSQPRGRQPIPKAR 69

RESULT 15
US-09-756-875-8
; Sequence 8, Application US/09756875
; Patent No. US20020150990A1
; GENERAL INFORMATION:
; APPLICANT: PIKE, IAN
; TITLE OF INVENTION: HEPATITIS C VIRUS PEPTIDES
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESS: Suite 701-E Columbia Square
; STREET: 555 13th Street, N. W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: U. S.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/756,875
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: US 08/259,721
; FILING DATE: 29-AUG-1994
; APPLICATION NUMBER: PCT/GB93/00410
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: ERNST, BARBARA G.
; REGISTRATION NUMBER: 30,377
; REFERENCE/DOCKET NUMBER: 1808-157A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)783-6040
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 97 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-09-756-875-8

Query Match 25.7%; Score 49; DB 9; Length 97;
Best Local Similarity 100.0%; Pred. No. 9.5e-37;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 DVKFPGGQIVGGVLLPRRGRLGVRATRKTSERSQPRGRQPIPKAR 69
Db 21 DVKFPGGQIVGGVLLPRRGRLGVRATRKTSERSQPRGRQPIPKAR 69

Search completed: October 30, 2004, 02:03:53
Job time : 50 secs
```

